

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2003, 03:14:59 ; Search time 2282 Seconds

(without alignments)  
13620.419 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068

Sequence: 1 atgtccatcacaacaaagac.....gtgtctgagaccatccctt 1068

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : GenEmbl:\*

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_man:\*  
37: em\_hcg\_vtl:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.2	5.1	7218	6 166494	166494 Sequence 14
2	52.6	4.9	1141	6 AX083744	AX083744 Sequence
3	50	4.7	4197	3 AF153835	AF153835 Dictyostel
4	48.4	4.5	441	1 RP282600	282600 R. prowazeki
5	47.8	4.5	9829	9 AL353689	AL353689 Human DNA
6	46.6	4.4	146568	2 AC022651	AC022651 Homo sapi
7	46.6	4.4	165236	9 AC024255	AC024255 Homo sapi
8	46	4.3	1141	6 AX083744	AX083744 Sequence
9	46	4.3	174253	2 AC068590	AC068590 Homo sapi
10	46	4.3	187254	2 AC124649	AC124649 Homo sapi
11	46	4.3	195912	2 AC103465	AC103465 Rattus no
12	45.4	4.3	114741	2 AC124954	AC124954 Medicago
13	44.6	4.2	282610	1 RPXX01	AJ235270 Rickettsi
14	44.6	4.2	304050	1 AP004829	AP004829 Scaphyloc
15	44.6	4.2	307750	1 AP003136	AP003136 Scaphyloc
16	44.4	4.2	138016	9 CNS01DTP	AL132827 Human chr
17	44.4	4.2	171942	2 AC021715	AC021715 Homo sapi
18	44.4	4.2	175948	2 AC019192	AC019192 Homo sapi
19	44.2	4.1	77096	2 AC116030	AC116030 dictyoste
20	44.2	4.1	148566	2 AC097671	AC097671 Rattus no
21	44.2	4.1	181254	2 AP000825	AP000825 Homo sapi
22	44.2	4.1	181589	9 AP000868	AP000868 Homo sapi
23	44	4.1	19089	2 AC026872	AC026872 Homo sapi
24	44	4.1	174603	9 AP003096	AP003096 Homo sapi
25	44	4.1	188250	9 AP001788	AP001788 Homo sapi
26	43.8	4.1	14461	6 AX256438	AX256438 Sequence
27	43.4	4.1	146383	2 AC116367	AC116367 Oryza sat
28	43.4	4.1	168698	9 AC068138	AC068138 Homo sapi
29	43.2	4.0	29520	3 CEM06F12	283244 Caenorhabdi
30	43.2	4.0	97683	2 AC116548	AC116548 Dictyoste
31	43.2	4.0	104025	9 AL138679	AL138679 Human DNA
32	43.2	4.0	104485	2 AC097864	AC097864 Rattus no
33	43.2	4.0	158840	9 AC091577	AC091577 Homo sapi
34	43.2	4.0	165018	9 AC079173	AC079173 Homo sapi
35	43.2	4.0	172657	9 AC015963	AC015963 Homo sapi
36	43.2	4.0	189148	2 AC068262	AC068262 Homo sapi
37	43	4.0	110000	2 AC113004_0	AC113004 Mus muscu
38	43	4.0	122515	2 AC120261	AC120261 Rattus no
39	43	4.0	348650	1 AP003364	AP003364 Scaphyloc
40	42.8	4.0	5201	1 AP011104	AJ011104 Apple pro
41	42.8	4.0	73039	2 AC116957	AC116957 Dictyoste
42	42.8	4.0	110000	2 AC125102_2	Continuation (3 of
43	42.8	4.0	124191	9 AC005153	AC005153 Homo sapi
44	42.8	4.0	162925	2 AP004461	AP004461 Oryza sat
45	42.8	4.0	175905	2 AC127263	AC127263 Mus muscu

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
166494	166494	Sequence 14 from patent US 5670367.	166494	1	GI:2724471	Unknown.	Unknown.	1 (bases 1 to 7218)	Dorner, F., Scheitlinger, F. and Falkner, F. Gunter.	Recombinant fowlpox virus	US 5670367-A 14 23-SEP-1997;	Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

```

source          1..7218      /organism="unknown"
BASE COUNT     1944 a    1491 c    1486 g    1929 t    368 others
ORIGIN
Query Match           5.1%; Score 54.2; DB 6; Length 7218;
Best Local Similarity 14.5%; Pred. No. 0.38;
Matches   83; Conservative 220; Mismatches 268; Indels 0; Gaps 0;

Qy  142 ACCGAATTGATTTGTACAGTATGCTAGTCCTTCAAAAGCTCCAGATGTTTCCAACT 201
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1021 ACAGAATTATATCCGAGCTTGCGTAGCTCAGAGGAGCTTCGATTTTTTTTTTTT 1080
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  202 GATGACTTTCCCTCATATCTGTGTTCTTTGGGCGGTGCAAGAGTTGGATGATCA 261
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1081 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1140
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  262 TTGGTTCGTGGTTACTGATGATTTGCATAGTTTGTTCCTCAAAGTCCCTGTCAAT 321
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1141 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1200
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  322 GGTTCCGTTAAGTTCCCAACAATCTGTCTCAAACCTTTTAATGTCTCACCAAAAT 381
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1201 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1260
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  382 GGNACTCAACAAGCATCTCCCTTTAGAAATGGCTCAAAAGGTGGTATGAACAAAT 441
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1261 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1320
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  442 GTTTATCCAGATGTTGCCCTCTTCTAGTTCTTTACAGTTTTCGATTTGATCAACAATTA 501
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1321 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1380
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  502 CTCGATCATCATCATCAGCTGAGTGTATATCAAGCGCTCGATCTTCCAAATCTGTT 561
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1381 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1440
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  562 GACCAAGTCAACAGAGATATCACTCAAAATATAGAACCATTTGGATTCAACAGTTGTT 621
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1441 AATTCTTATCTCTTTAATCTACTTGCAATAGATAGTATATACAGATGAGCTTACATGCG 1500
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  622 GCCTCTCAAAGAGATATAATTAATCTGTGPAAGCAAGTAAAACAATTTCAAACTACTAT 681
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1501 GTTTTGTGAAACGAAATAGATGCTGTGAAGAGATGCTACGCTAGTCAATATCACCCT 1560
               : : : : : : : : : : : : : : : : : : : : : : : :
Qy  682 GTCCGATATAGTGAAGTANGTGTGAATTA 712
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1561 TTCATATTTAGAAATATATGATGTAAAAATA 1591
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
AX083744 LOCUS 1141 bp DNA linear PAT 28-FEB-2001
DEFINITION Sequence 22 from Patent WO0111061.
VERSION AX083744
KEYWORDS AX083744.1 GI:13185472
SOURCE .
ORGANISM synthetic construct.
REFERENCE artificial sequences.
AUTHORS 1 (bases 1 to 1141)
TITLE Kunst,L. and Clemens,S.
JOURNAL Regulation of embryonic transcription in plants
PATENT Patent: WO 011061-A 22 15-FEB-2001;
UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES
source Location/Qualifiers
promoter 1..1141
/promoterm="synthetic construct"
/db_xref="taxon:32630"
1..1141
/note="consensus sequence of A.t., L.a., and B.n. PAEI
promoters"

```

```

ORIGIN
Query Match          4.9%; Score 52.6; DB 6; Length 1141;
Best Local Similarity 11.0%; Pred. No. 1.1;
Matches 53; Conservative 196; Mismatches 234; Indels 0; Gaps 0;

OY      43 TATCTTCCTTGTGGAAATGAAAAAGCGCAACCAACTTAATTGATCAATTTGATCAAGGAT 102
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       182 HMMNNNGEBITWVRYKTDHDSBKNNNYGMWMKKMSYVTYYWWMDMCKRKVRW 241
OY      103 GTTTTGCCAACTCAGGTTCCGGAATATAATATGCAGAATATACCAGATTGATTGTTACAGT 162
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       242 VTRGRMRNVAVVAABTAIRRRYRNNGBAMAYRRWTNNNNNNNAKMKCRAKYGMWRAB 301
OY      163 GATGCTAGCTTCAAAGCTCCAGATGGTTTTCTCAACTGATAGCATTTTCTTCATAT 222
        ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       302 VNSTGTWKSKTKTVKRSICMANNCRAADANKDHKKMKKSALMGYYMNNNNNNNTYKKAR 361
OY      223 CTGTGTTCTTTGGGCGTGCGTCAAGATTGGATGATCATTTGCTGCGTGTACTGGT 282
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       362 HBARDVMWHSAWKMHANAHYERKKMTBYKRKTVMNNTGGTTMWKMYAWYMKMDWD 421
OY      283 GATTTCGATGATTGTTGTTTCTCTCAAGGCCCTGTCAATGTTCCGTTTATGTTGCCA 342
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       422 BGYVNNNNNGRIYGYMKNKMMTYTMKANNCKWARMHDKTHHNNTTWKKTKTYWN 481
OY      343 CAATCTGTGCTCAAACTTTTATTTATGTGCTCACCAATGCTCAACAAGCATCTTCC 402
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       482 CYWSMTGSKSHRBAAAYTTYMMWRHYAHANNNDYMKACTKYIVCSKMNNYA 541
OY      403 CTTTAGAATTGGCTCAAAAAGTTGGTATGACAAATGTTTATTCAGATGTGCTCT 462
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       542 AMYTSSNNYISRYRMKTNNSWMBSDTSRGMRANNYARABHYGYKNTRWMBESHTB 601
OY      463 TCTAGTTTTCACAGTTTTCGATGTGATTAACAATTAATCTCCAATCATCATCAGCT 522
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       602 HBRAGAATHYMBMYBAKHCMKAKYAKKYAAGGSNNNNNNNNNNNNNNNNNNATCARD 661
OY      523 GCA 525
        :|
Db       662 YYA 664

RESULT 3
AFI63835 4197 bp mRNA linear INV 01-MAR-2000
LOCUS AFI63835
DEFINITION Dictyostelium discoideum cellulose synthase (dcsa) mRNA, complete cds.
ACCESSION AFI63835
VERSION AFI63835.1 GI:602920
KEYWORDS .
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS 1 (bases 1 to 4197)
TITLE Blanton,R.L., Fuller,D., Iranfar,N., Grimson,M.J. and Loomis,W.F.
JOURNAL The cellulose synthase gene of Dictyostelium
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (5), 2391-2396 (2000)
PUBMED 20160962
10681463
2 (bases 1 to 4197)
REFERENCE Blanton,R.L., Fuller,D., Iranfar,N., Grimson,M.J. and Loomis,W.F.
AUTHORS Direct Submission
JOURNAL Submitted (29-JUN-1999) Genar for Molecular Genetics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES
source Location/Qualifiers
1..4197
/oranism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
1..4197
/gene="dcsa"
806..3985

```

```

/gene="dcsA"
/notes="DcsA"
/codon_start=1
/product="cellulose synthase"
/protein_id="AAF0200.1"
/db_xref="GI:6002921"
/translation="MDNRGGDPPIINTPNINSSGSGSYNNSSNNIGRIDGNNOSS
GRNLKPKPSOSNLKWIARDLKKKSVKRDSEKLKSGYLKKATVMPGEDGSGDD
GNTEGLPISEGMDLPPSSNSRGSGSNDGKKOFPYKEMNSPSEYGTITSGQGFDTL
VDPISLAMEKKRKHQVYQEOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO
VAKDSETLTELIIGEEKKEVVEVYKPFESH1LQATPAVFLIIM1PFAAYAGWTNR
TDYITFSYLIETIIEFISFLGSAHLNFPNCTPVALVLEOILKRRKQPTVMR
VYVCYKEPSSIVSRTPRTAISMDYPSMLTIGLLDSSYVESGSAHLOSVEKPIY
VLQKATYSVHNIIRPPTVSGHDEPHGILNETSKSTESKTEVIAEYQVETETPLNS
WFGGQELPRADDAERLIALKLDNDNESPVRTTKSESEKISNFT1DS1QSLMHGSA
PFRLPISILTKDQYVNFVSELNNOHRLERLNETALMAQYUQLMAGDELPDELSI
SGNRIPIDTCDGIVSPKCYLRRRKP1PHNKAQINNALPNSTKADYEFGLJLD
ADQOPHPFLKRVLPYFSGDEGDLAPVOTQPPFSNIYVDDPLGHNMMEFYGVMSG
RSANACPFCVGTNALIFRQPLYDIGILNMSVTEDMTGMLQVYSKSYHNEVLVYV
GTAVDLKETLEQGRMAQGAVEIFSLTPWGYINGKLGWRMLYNLDSCTPIFSLPFA
FYFASPLIMASTVNIWPIVADPILFILAQM1PVVLIQYMLIRAKRPYEAKRSDP
SLWEATDLMAEQOTFFGACTYIISNMEGSA1LILKAKIRSHLALMMNRKADPVC
KKPVCVEYRQTKLVNENDNAOESGKKAQOSRSTSKESDITKNSRLPLPNI1LPEV
VILLAMSAVLRPNCFQNDMLLVYVAGFSSTLHMLSTFIPALRROSEKQVPAASSY
HAHNIIVFLVIGFLVLEFVDVYKVCIPRG"

```

Query Match	4.7%	Score 50;	DB 3;	Length 4197;
Best Local Similarity	45.8%;	Pred. No. 2.9;		
Matches 209;	Conservative	0;	Mismatches 245;	Indels 2;
				Gaps 1;

Oy	429	TTATACACAAATGTGTTATTCAGATGTGCCCTCTCTAGTCTTTACAGTTTGCGATT	488
Db	181	TTATATACACAACTATTTTAAATTTTATTTAATTTATTTATTTATTTATTTT	240
Oy	489	GTAACACAAATTACTCCAAATCATCATCAGCTGCAGTGTGATCAAGCCCTGATCT	548
Db	241	TTTTTAAATAATTTATTTATTTATCTTTTATTTTCTTTTTTTTATTTATTTAATTTA	300
Oy	549	TCCACAATCTGGTGCCCAAGTCAACAGATTTCACATAAATATAGAACCATTTTGA	608
Db	301	TAAATATTTAAACAACAACAACAACAACAACAACAACAACAACAAGTTTTTA	360
Oy	609	TTCAACAGTGTGTGCTCTCAAGAGAAATATATTAACTCTGTAAAGCAAGTAAACCAAT	668
Db	361	CTTTATATTTTTTATTTATGTGTGA - ATTAAGAAATTTAAACTAATTAATTTAATTTGTA	418
Oy	669	TTCAAACTATCTATGTGCGATATATGTGAAGAATATGTGTGAATTTAAGATATCATCAGGA	728
Db	419	ATACCATATCCATATAAATAATTTATTTATTTTGGGTAAATTAATTAATTAATTAACATAAA	478
Oy	729	TCAACATATCAATGTTCACATCATTTGGTGCCTGTGATTAAGCATCGTTTATACGATGT	788
Db	479	TCTCCAAAAATCTTTAAATCAAAAAAAATTAATATATATATATATATTTATGTGTTTT	538
Oy	789	TTTGCTTTGAATTCCAATTGTGTGATGAAGAAAGCAAAAGTGTGCTGTGAAGTTATCA	848
Db	539	GTGTGTTTAAATTAATATTTGTTGTGTCAATCAATTTATTTTAAAAAATAAAAA	598
Oy	849	GAATTTATGACTATATCTTTGTTGGACTGTT 884	
Db	599	AAAAAAAAATATATGTGTGTTTTTTTTTTTTTT 634	

RS0014					
RP282600/c					
LOCUS	RP282600	441 bp	DNA	linear	BCT 07-MAR-1997
DEFINITION	R. prowazekii genomic DNA fragment (clone A340R).				
ACCESSION	Z82600				
VERSION	Z82600.1	GI:1871951			
KEYWORDS	Rickettsia prowazekii.				
SOURCE	Rickettsia prowazekii.				

ORGANISM	Rickettsia prowazekii
REFERENCE	Bacteria: Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiaceae; Rickettsia; typhus group. 1 (bases 1 to 441)
AUTHORS	Anderson,S.G., Eriksson,A.S., Naslund,A.K., Andersen,M.S. and Kurland,C.G.
TITLE	The Rickettsia prowazekii genome: a random sequence analysis
JOURNAL	Microb. Comp. Genomics 1 (4), 293-315 (1996)
MEDLINE	98353639
PUBMED	9689214
REFERENCE	2 (bases 1 to 441)
AUTHORS	Anderson,S.G.E.
TITLE	Direct Submission
JOURNAL	Submitted (31-OCT-1996) Anderson S G E., Department of Molecular Biology, Husegatan 3, BOX 590, Biomedical Center, S-751 24 Uppsala, SWEDEN
COMMENT	The overall error rate is an estimated 0.5%.
FEATURES	Location/Qualifiers

BASE COUNT	191 a	57 c	37 g	155 t	1 others
ORIGIN					

Query Match	4.5%	Score	48.4	DB	1	Length	441
Best Local Similarity	47.1%	Pred. No.	9.4				
Matches	145	Conservative	1	Mismatches	162	Indels	0
						Gaps	0

Qy	667	ATTTCAACTCTAGTGGGATATGTGAAGTAGTGTGAAATTAAAGGATATCATCACA	726
Db	420	ATTTCAACGAAAAATTAAGTAGATTAATGCTCTTATCAATTAGATTACATTAT	361
Qy	727	GATCAACATACAAATGTTCAACTCATTTGGTACCTCTGATTAAGCCATACGTTATACTGAT	786
Db	360	TATATTCATTAATTAAATCCATATATTAATATTTTATATAGAACAAATTTTGAATTGG	301
Qy	787	GTTTTCGCTTGAATTCCAATTTGTGTGATGAAAAAGCAAAAGTTGCTGTGAAGTTATC	846
Db	300	AATTTTATTAATTAGATTAATTGCTTTTAAAAAGCACAABAAGCATTAATATCTGTGTAGAA	241
Qy	847	AAGAAATTATTGACTATAACTTTAGTTTGCACITTTGGGCTCTCGGATTAACCTCCCA	906
Db	240	GATTAATTTTGTAGTAAATTAATTAATGATGTGTTATTAATTCGCAATGGGTTAACCGTTCA	181
Qy	907	GCCACACGAAGTGTATATGCTCATTTGGCTAAATATATCAAACTTTTATGCTCAATTGAGC	966
Db	180	GTTCTTATTATAGGTATAGTGTCTATGCTATAGGAGGAAATTTGATTAAGAATTTGAGC	121
Qy	967	CAACAATT 974	
Db	120	TTTAAACT 113	

RESULT 5	ALJ53689	195829 bp	DNA	linear	PRI 21-OCT-2001
LOCUS	ALJ53689				
DEFINITION	Human DNA sequence from clone RP5-1087E8 on chromosome 19q2.11-42.2, complete sequence.				
ACCESSION	ALJ53689				
VERSION	ALJ53689.26	GI:16304897			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 195829)				
AUTHORS	Moore, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,				



\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved. 2283: contig of 2283 bp in length  
1  
2284 2383: gap of 100 bp  
2384 9073: contig of 6690 bp in length  
9074 9173: gap of 100 bp  
9174 16904: contig of 7731 bp in length  
16905 17004: gap of 100 bp  
17005 25601: contig of 8597 bp in length  
25602 25701: gap of 100 bp  
25702 34479: contig of 8778 bp in length  
34480 34579: gap of 100 bp  
34580 41772: contig of 7193 bp in length  
41773 41872: gap of 100 bp  
41873 55432: contig of 13560 bp in length  
55433 55532: gap of 100 bp  
55533 68980: contig of 13448 bp in length  
68981 69080: gap of 100 bp  
69081 84607: contig of 15527 bp in length  
84608 84707: gap of 100 bp  
84708 100925: contig of 16218 bp in length  
100926 101025: gap of 100 bp  
101026 115546: contig of 14521 bp in length  
115547 115646: gap of 100 bp  
115647 146568: contig of 30922 bp in length.  
Location/Qualifiers  
1. 146568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-28H7"  
/clone\_1ib="RP11-11 Human Male BAC"

## FEATURES

source  
misc\_feature  
1. 2283  
/note="assembly\_fragment"  
2384. 9073  
/note="assembly\_fragment"  
misc\_feature  
9174. 16904  
/note="assembly\_fragment"  
misc\_feature  
17005. 25601  
/note="assembly\_fragment"  
misc\_feature  
25702. 34479  
/note="assembly\_fragment"  
misc\_feature  
34580. 41772  
/note="assembly\_fragment"  
misc\_feature  
41873. 55432  
/note="assembly\_fragment"  
misc\_feature  
55533. 68980  
/note="assembly\_fragment"  
misc\_feature  
69081. 84607  
/note="assembly\_fragment"  
misc\_feature  
84708. 100925  
/note="assembly\_fragment"  
misc\_feature  
101026. 115546  
/note="assembly\_fragment"  
misc\_feature  
115647. 146568  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_end:left"  
BASE COUNT 44763 a 27525 c 28254 g 44924 t 1102 others  
ORIGIN

Query Match 4.4%; Score 46.6; DB 2; Length 146568;  
Best Local Similarity 48.0%; Pred. No. 7.1;  
Matches 133; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 584 CTCAAAATATAGAACCATTTGGATTCAACAGTGTGGCTCTCAAGAGATATATTA 643  
Db 118836 CACGAAATGCACATCAAACTCAATGAGATATCATCTCAACCCCACTTAAATGCTTA 118895

QY 644 ACTCTGTAAAGCAAGGTAAACCAATTTCAACTACTATCTGCGATATAGTAAAGATGT 703  
Db 118896 TATCCAAAGACAGCGCAATACAAATGCTGGTGAATGTGGTAAAGGAAACCTCGT 118955  
QY 704 GTGAATTAAGGATTCATCAGAGATCAGATCAACATATCAATGTTCACATCTTGACTCTG 763  
Db 118956 ACACATTAAGTGAAGATGTAAATTAAGACACAACTACATGAGATCTTTGACAGTTCTT 119015  
QY 764 ATAGGCATACGTTTATACGATGTTTGGCTTGAATTCATTTGCTGATGAGAAAGC 823  
Db 119016 CAAAACTTAAATACCTGTGATATGATGCGAGAAATCCACTGCTGGATATATACC 119075  
QY 824 AAAAGTGTCTGTGAAGTTATCAAGATTTATTTGAC 860  
Db 119076 AAAAGAAAGAAATTCAGTGTATCAAGAGATCTTAC 119112

RESULT 7  
AC024255/c  
LOCUS  
DEFINITION  
BAC library complete sequence.  
AC024255  
VERSION  
AC024255.22 GI:11465004  
HTG.  
SOURCE  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 165236)

## REFERENCE

## AUTHORS

Muzny,D.M., Adams,C.C., Adio-Oduola,B., Alt-omman,F.R., Allen,C.,  
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbieri,J.,  
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Buck,J.,  
Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,  
Burich,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,  
Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,  
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Demn,A.L., Ding,Y., Dim,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,  
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
Guaracane,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F.,  
Howard,S., Huber,J., Huylk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsone,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
LouiSeged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martini,R.,  
Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,  
Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,  
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,N.,  
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,M.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,B., Nwokwenta,S.,  
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojase,A., Rojupokan,I., Rolfe,M.,  
Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shim,C.,  
Shooshartai,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A.,  
Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamezia,A.,  
Tameria,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,  
Thomas,N., Thomas,S., Usmani,K., Vazquez,L., Vera,V., Villalon,D.,  
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
Washington,C., Watlington,S., Williams,G., Williamson,A.,  
Wleszyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
Zorilla,S., Kucherlapati,R. and Gibbs,R.  
Direct Submission

## TITLE

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 165236)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (28-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 165236)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (30-NOV-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 30, 2000 this sequence version replaced gi:1125287.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and CDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig Length: 165236  
Phrap values in estimate: 164180  
Average error rate (BCM-Phrap estimate): 0.00021131  
Fraction of Phrap values less than 40 : 0.02426  
Number of consensus changing edits : 26  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context  
4248 gtgcaggcac (n) acaaaaaaag gtgcaggcac (a) acaaaaaaag  
25468 tttaagatc (n) atctctctc tttaagatc (t) atctctctc  
35515 tgcctcacg (n) ccaattctc tgcctcacg (t) ccaattctc  
35635 acatctcga (n) tactgggac acatctcga (c) tactgggac  
36329 cagccctggc (n) acatggtgaa cagccctggc (a) acatggtgaa  
42433 ttctgttgg (n) tttaaaaaaa ttctgttgg (t) tttaaaaaaa  
52516 ttaattatg (n) catcatctg ttaattatg (t) catcatctg  
59538 caaagcaagt (n) acatgggaact caaagcaagt (t) acatgggaact

90839 actgatcac (n) ngaaagatc actgatcac (a) agaaagatc  
90840 ctgatcacn (n) gaaagatcga ctgatcacn (g) aagatcga  
90843 gatacanga (n) aagatcga gatacanga (g) aagatcga  
91026 caacacttg (c) gggacattc caacacttg (t) gggacattc  
101466 catcaattg (n) tcaacaata catcaattg (t) tcaacaata  
101608 aaataaac (n) ccgtrtatc aaataaac (c) tgaatggcc  
109374 cctccacct (t) tgaatggcc cctccacct (c) tgaatggcc  
109693 accttatgg (n) agaatgatc accttatgg (t) agaatgatc  
109738 ctgggtcaaa (n) ggtatctcg ctgggtcaaa (t) ggtatctcg  
118157 aatcttaaa (n) tcaatgaacat aatcttaaa (t) tcaatgaacat  
139324 ggaacttagt (n) aannatttc ggaacttagt (g) aannatttc  
139327 cttagtgaa (n) mactttctc cttagtgaa (t) mactttctc  
139328 cttagtgaa (n) mactttctc cttagtgaa (t) mactttctc  
139329 ttatgnaann (n) attttcttc ttatgnaann (t) attttcttc  
139330 ttatgnaann (a) attttcttc ttatgnaann (t) attttcttc  
146052 ccttaaac (t) acaaggcct ccttaaac (g) acaaggcct  
146207 gtcatcagg (n) aatggaatt gtcatcagg (a) aatggaatt  
160897 accctcgag (n) tttagagta accctcgag (c) tttagagta

## ----- Distribution of Quality &lt; 40 Bases -----

#	5	10	15	20	25	30	35	40
1000							*	*
900							*	*
800							*	*
700							*	*
600							*	*
500							*	*
400							*	*
300							*	*
200							*	*
100							*	*
0	*	*	*	*	*	*	*	*

Phrap Value Range

Version: 1.01 gxf0.  
FEATURES Location/Qualifiers  
source 1..165236 Homo sapiens  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-31A23"  
2..309  
repeat\_region /rpt\_family="L1P"  
complement(312..1362)  
repeat\_region /rpt\_family="L1M4"  
complement(1368..1821)  
repeat\_region /rpt\_family="L1P4S"  
1822..2720  
repeat\_region /rpt\_family="L1P2"  
2705..5433  
repeat\_region /rpt\_family="L1PA12"  
5434..5788  
repeat\_region /rpt\_family="THE1B"  
5789..6973  
repeat\_region /rpt\_family="L1PA12"  
6974..7278  
repeat\_region /rpt\_family="A1U5p"  
7279..7425  
repeat\_region /rpt\_family="L1PA12"  
7428..7874  
repeat\_region /rpt\_family="L1"  
7876..7959  
repeat\_region /rpt\_family="L1MCc"  
7960..7979  
repeat\_region /rpt\_family="TAA)n"  
7980..8265  
repeat\_region /rpt\_family="L1MCc"  
complement(8266..8568)



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheeters,R., Meldrum,J., Menus,L., Mihov,T., Miranda,C., O'Connor,T., O'Donnell,P., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,U., Peterson,K., Pierre,N., Pisaní,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Teste,S., Theodore,J., Turrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 174253)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Baetsen,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Hand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marcus,R., McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheeters,R., Meldrum,J., Menus,L., Mihov,T., Miranda,C., Mlenga,V., Morrow,D., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,U., Peterson,K., Pierre,N., Pisaní,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Teste,S., Theodore,J., Turrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 9, 2000 this sequence version replaced gi:17705204.  
All repeats were identified using RepeatMasker:  
Salt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L9496  
Center clone name: 131 K16

----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156518 bases at least Q40  
Consensus quality: 165902 bases at least Q30  
Consensus quality: 169236 bases at least Q20  
Insert size: 170000; agarose-fp  
Insert size: 171353; sum-of-contigs  
Quality coverage: 3.5 in Q20 bases; agarose-fp  
Quality coverage: 3.5 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequences as soon as it is available and the accession number will be preserved.

FEATURES	source
1	1232: contig of 1232 bp in length
1233	1332: gap of 100 bp
1333	3334: contig of 2002 bp in length
3335	3434: gap of 100 bp
3435	5627: contig of 2133 bp in length
5628	5727: gap of 100 bp
5728	7753: contig of 2026 bp in length
7754	7853: gap of 100 bp
7854	9492: contig of 1639 bp in length
9493	9592: gap of 100 bp
9593	11429: contig of 1837 bp in length
11430	11529: gap of 100 bp
11530	13178: contig of 1649 bp in length
13179	13278: gap of 100 bp
13279	15953: contig of 2675 bp in length
15954	16053: gap of 100 bp
16054	18335: contig of 2282 bp in length
18336	18435: gap of 100 bp
18436	19919: contig of 1484 bp in length
19920	20019: gap of 100 bp
20020	22351: contig of 2332 bp in length
22352	22451: gap of 100 bp
22452	25473: contig of 3022 bp in length
25474	25573: gap of 100 bp
25574	28192: contig of 2619 bp in length
28193	28292: gap of 100 bp
28293	32574: contig of 4282 bp in length
32575	32674: gap of 100 bp
32675	36983: contig of 4309 bp in length
36984	37083: gap of 100 bp
37084	41557: contig of 4474 bp in length
41558	41657: gap of 100 bp
41658	46775: contig of 5118 bp in length
46776	46875: gap of 100 bp
46876	52390: contig of 5515 bp in length
52391	52490: gap of 100 bp
52491	57638: contig of 5148 bp in length
57639	57738: gap of 100 bp
57739	65255: contig of 7517 bp in length
65256	65355: gap of 100 bp
65356	71907: contig of 6552 bp in length
71908	72007: gap of 100 bp
72008	79155: contig of 7148 bp in length
79156	79255: gap of 100 bp
79256	88577: contig of 9322 bp in length
88578	88677: gap of 100 bp
88678	98500: contig of 9823 bp in length
98501	98600: gap of 100 bp
98601	106888: contig of 8288 bp in length
106889	106988: gap of 100 bp
106989	117423: contig of 10435 bp in length
117424	117523: gap of 100 bp
117524	125991: contig of 8466 bp in length
125992	126091: gap of 100 bp
126092	137392: contig of 11301 bp in length
137393	137492: gap of 100 bp
137493	154901: contig of 17409 bp in length
154902	155001: gap of 100 bp
155002	174253: contig of 19252 bp in length.
Location/Qualifiers	
1. 174253	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="8"	
/map="8"	
/clone="RP11-131K16"	
/clone_lib="RP11 Human Male BAC"	
1. 1232	
/note="assembly_fragment"	
1333. 3334	
/note="assembly_fragment"	
3435. 5627	
/note="assembly_fragment"	
misc_feature	
misc_feature	
misc_feature	



```

misc_feature      5728..7753
                    /note="assembly_fragment"
misc_feature      7854..9492
                    /note="assembly_fragment"
misc_feature      9593..11429
                    /note="assembly_fragment"
misc_feature      11530..13178
                    /note="assembly_fragment"
misc_feature      13279..15953
                    /note="assembly_fragment"
misc_feature      16054..18335
                    /note="assembly_fragment"
misc_feature      18436..19919
                    /note="assembly_fragment"
                    clone_end=17
                    vector_side:right"
misc_feature      20020..22351
                    /note="assembly_fragment"
misc_feature      22452..25473
                    /note="assembly_fragment"
misc_feature      25574..28192
                    /note="assembly_fragment"
misc_feature      28293..32574
                    /note="assembly_fragment"
misc_feature      32675..36983
                    /note="assembly_fragment"
misc_feature      37084..41557
                    /note="assembly_fragment"
misc_feature      41658..46775
                    /note="assembly_fragment"
misc_feature      46876..52390
                    /note="assembly_fragment"

```

Query Match Best Local Similarity 4.3%; Score 46; DB 2; Length 174253;  
Matches 118; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

QY 482 TCGGATTGTATCAACAATTATCTCCATCATCATCATGCTGCGATGATATCAAGCCCT 541
Db 59082 TTGGAGAGTAATATCTAAAAAATCATTTGCTATTCGACGACATTCATTTGATATATT 59023
QY 542 CTGATCTTCCACATCTGCGACCAAGTCAACAAGATATCACTCAAAAAATATGAACCA 601
Db 59022 CAGAGACTCCCTTGTATGTGTGCAACTAAAGAACTGAACATGTATGTAGATGAG 58963
QY 602 TTGTGATTCACAGCTGTGCTCTCAAGAGAAATATTAATCTCTGTAAAGCAAGATA 661
Db 58962 GTTATATGACTACAGTTGCCATCTATATAGCTGTGATTTTATTTAAAAACCTGTCA 58903
QY 662 AACCAATTTCAACTACTATGTCGATATAGTGAAGATATGTGAATTAAGGATAT 719
Db 58902 AACCAATCTAATATAGCTGTGAGCTTCCCATTTGAATCGGAATTAAGAAATTT 58845

```

```

RESULT 10
AC124649/c 187254 bp DNA linear HTG 06-AUG-2002
LOCUS Homo sapiens chromosome 8 clone RP11-2318 map 8, *** SEQUENCING IN
DEFINITION
PROGRESS ***, 2 ordered pieces.
ACCESSION AC124649
VERSION AC124649.3 GI:22123145
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULFILL; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 187254)
Birten,B., Nusbbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-2318
Unpublished
2 (bases 1 to 187254)
Birten,B., Lincon,L., Nusbbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 187254)  
Birten,B., Nusbbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farrow,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Spencer,R., Strang-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 6, 2002 this sequence version replaced gi:21699345.  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L27405  
Center clone name: 23\_I\_8

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
1 64780: contig of 64780 bp in length  
64881 64880: gap of 100 bp  
64881 187254: contig of 122374 bp in length.  
Location/Qualifiers  
1..187254  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

FEATURES  
source





\* consists of 5 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1  
\* 3390 3389: contig of 3389 bp in length  
\* 3490 19779: contig of 16290 bp in length  
\* 19780 19879: gap of unknown length  
\* 19880 39080: contig of 19201 bp in length  
\* 39081 39180: gap of unknown length  
\* 39181 67046: contig of 27866 bp in length  
\* 67047 67146: gap of unknown length  
\* 67147 114741: contig of 47595 bp in length.  
Location/Qualifiers  
1. 114741  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone\_11b="Medicago truncatula BAC library H1"  
BASE COUNT 39037 a 19145 c 18697 g 37461 t 401 others  
ORIGIN  
Query Match 4.3%; Score 45.4; DB 2; Length 114741;  
Best Local Similarity 50.2%; Pred. No. 13;  
Matches 112; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 604 TTGATTCACAGTTGTCCTTCACAAAGATATATTAAGTCTGTAAGCAAGTAA 663  
Db 12845 TGGAAATCAGCAAAATTTATTTGATATATGAGATGTTAAATCAAAAGCAT 12786  
QY 664 CCAATTCACAACTGATGTCGATATGATGATGATGATGATGATGATGATGATG 723  
Db 12785 TGTCAAAAATTTATTTAGTTCTATGTTACCTTATGTTGATGATGATGATGAT 12726  
QY 724 AGAGATCAACATCAATGTTCAACTCATGTCCTGATTAAGCCATACGTTTACT 783  
Db 12725 TAACATCTTCAACATGATGTTCTGATTCAGTTATTTTATTTAAGAAATTCGATTATGCT 12666  
QY 784 GATGTTTGGCTTGAATTCCAATTTGTGTGATGAAAACCAA 826  
Db 12665 GTTGATTTTGTGTTGCTTTCGAAATGGCTAGTAATATGTCATA 12623  
RESULT 13  
R 13  
L 13  
DEFINITION REXX01 282610 bp DNA linear BCT 11-NOV-1998  
1/4.  
Rickettsia prowazekii strain Madrid E, complete genome; segment  
AJ235270 AJ235269  
VERSION AJ235270.1 GI:3860572  
KEYWORDS complete genome.  
SOURCE Rickettsia prowazekii.  
ORGANISM Rickettsia prowazekii  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.  
1 (bases 1 to 282610)  
Andersson, S.G., Zomorodipour, A., Andersson, J.O.,  
Scheritz-Ponten, T., Almark, U.C., Podowski, R.M., Naslund, A.K.,  
Eriksson, A.S., Winkler, H.H. and Kurland, C.G.  
The genome sequence of Rickettsia prowazekii and the origin of  
mitochondria  
JOURNAL Nature 396 (6707), 133-140 (1998)  
MEDLINE 99039499  
PUBMED 9823893  
REFERENCE 2 (bases 1 to 282610)  
AUTHORS Andersson, S.G.E.  
TITLE Direct Submission  
JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson,  
Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University

of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN  
FEATURES  
source  
Location/Qualifiers  
1. 282610  
/organism="Rickettsia prowazekii"  
/strain="Madrid E"  
/db\_xref="taxon:782"  
gene  
1. 822  
/gene="RP001"  
CDS  
1. 822  
/gene="RP001"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14474.1"  
/db\_xref="GI:3860573"  
/translation="MTKLIIHLVSDSSVOTAKHAANSALAOFTSIKQKLYHWPIMRNC  
ELLNEVLAKIESKHGIVLTIADQELRKLTFCEYELKPCISIVGKIHKENSFSGI  
EIAEKQNVYKFKDKYFDTLNAIDVAIRHDDCOMINLESSEDIILGPRTSYTSFV  
FLAYNGLRKANIPYVNCPPPIEDKIDQLVGVGVINRNRIETFEARINLIQINEN  
KSYTDENIVORECTEVRKICQNMWPEVDVSTRSLEETALIMRIYRNKKYHK"  
906. 1298  
/gene="RP002"  
CDS  
906. 1298  
/gene="RP002"  
/codon\_start=1  
/transl\_table=11  
/product="THIOREDOXIN (trxa)"  
/protein\_id="CA14475.1"  
/db\_xref="GI:3860574"  
/translation="MSCVNEITITLLEFSDNDINTTORIMVNVVSDSPKNEVLES  
PVMVDMAWMCPCCKKLPIIDPISKELODKKVKVLMKNIDENPKTSEYGISIPFIM  
LFRNGEQDKTKGLQCKNSLDMINKSI"  
1309. 2067  
/gene="RP003"  
CDS  
1309. 2067  
/gene="RP003"  
/codon\_start=1  
/transl\_table=11  
/product="O-ANTIGEN EXPORT SYSTEM ATP-BINDING PROTEIN RPEB  
rfbB"  
/protein\_id="CA14476.1"  
/db\_xref="GI:3860575"  
/translation="MLNHSKVFIETDGVGSDIVHKAQCKHFKFLKKGVSLPTI  
PLINLNFSGYEGEKIAFGNSGSKSLKLKLAGIYPLKSGIVKHGIIAATIDGV  
GGEQGTGRANIKMLVNNMLDKYSKIEKIIPESEIGSKIDLEIKYSGMSRL  
AESVSIQNPQIILLDEVPAAGSYIEKSLNLMKKKFNFTPIISIIHQEBELIKDNC  
DRCILKDGHIIDGTPSEIFKIYQKSNKEIHK"  
2064. 2840  
/gene="RP004"  
CDS  
2064. 2840  
/gene="RP004"  
/codon\_start=1  
/transl\_table=11  
/product="O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RPEA  
rfbA"  
/protein\_id="CA14477.1"  
/db\_xref="GI:3860576"  
/translation="MIKYFSPSKYKWRALALVKAIVRONKDSFLGSLMSLIQPIYHI  
IYISYFPGFLRQPREFMVNLVGGIPIKSFVSLTICASSIVTRDOIKKVRISKY  
FPPIASLSQVLTILCSFSANYPATILPEKPSMGIIMPTLIIPIVIVCSGISAV  
AFETPIRIRPQMLNLVGLVYISIPVPIYSLIPSKKIYEFNFFLVIRPVQVLV  
INGSLPDMWLIIRKSVAVATVATVSIYSLIYRFSKVIYVL"  
2837. 3724  
/gene="RP005"  
CDS  
2837. 3724  
/gene="RP005"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14478.1"  
/db\_xref="GI:3860577"  
/translation="KLGFNLDLPNEIDLTLGLKDOIPLDIYLFKADSKLHKDMLMFLRA  
TPLSLIPDYSKFLKISPHLDPLAELFCISKVETISRLKHKDPIIYECRKKVQR

[illegible]

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
4.28; Score 44.6; DB 1; Length 282610;	47.9%; Pred. No. 16;	0;	0;	139;	0;	0;
708	AATTAAGGATTCATCGAGATCAACAATTAAGTTCACATCTAGTCTGATTA	767				
254837	AATTAGATTATTAACATTTATTAATTCATTAATTAATTCATATATATATTTTATATA	254896				
768	GCCATACGTTTATCTGATGTTTGGCTTGAATTCATTTGTGTGATGAAAGCAAAA	827				
254897	GAAGAAATTTTGAATTTGATTTTATTTATTTAGATTAATGCTTTTAAAGCAGAAAAG	254956				
828	GGTTCCTGTTGAAGTTATCAAGAAATTTATTTAGCTTAATCTTTAGTTTGGACTTTGGG	887				
254957	CATATAATCTTGTTGAGAAATTTATTTTATATTAATTAATGATGTTGTTATTAATTTGC	255016				
888	TCTCGGATTAATCTCTCCAGCCACAGAAATGATATTTGCTCAATTTGGTAAATCATCAA	947				
255017	CATTGGGTTAACCGTTTCAGTTCTTATTTAGGATATGTCATGCTATAGAGAGAAA	255076				
948	CTTTATGCTCAATTTGAGCCACAATTT	974				
255077	ATTTCATTAAGAAATTTGAGCTTAATAACT	255103				

complement (9267. .9339)  
/gene="tRNA\_Phe (GAA)"  
9661. .10686  
/gene="RP011"  
9661. .10686  
/gene="RP011"  
/codon\_start=1  
/transl\_table=11  
/product="NIFR3-LIKE PROTEIN (nifR3)"  
/protein\_id="CA14483.1"  
/db\_xref="GI:3860582"  
/translation="MLNYDIMLNFITFMKIGNIELSNVLIAPMSDVTDLERKLV  
KRGAGVSEMIASRAMIKSRQSMOKCAIMHDDPSACVOLAGSPDIAEAKGN  
EDWAKIIDNIFGCPAKVYVAGSALRMCDEDLAAKIPATYKAVKIPATVKRKN  
DDNTRAPLTLAKIAPNSGVOMTVHGRTRCQFISGNANMDFITVEAVKIPYANSD  
ITNFAKKEALQSGADGIWVGWYGEKPLISQIAYLTGKEKAPVIAEODIIT  
KHTDAIIDYGGKSVGPVIAKHIWYSGNGLSPSSAEFRCTVLMKMDPIAVEKIAIEFYM  
SNDANK"

10683. .11756  
/gene="RP012"  
10683. .11756  
/gene="RP012"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14484.1"  
/db\_xref="GI:3860583"  
/translation="MKRLSFFIIFLFNNSVAVEKAAIITDYKPIFLPVITENKKIK  
IAIRSYLKNKSYFVLVDPSFTEIYLOELVILPARKIEKNLLIKNTSYIKALN  
KNPIDSLOSFPKODESRNKTDTNTIYEHKLNKSKSHSSTPMNSTVQOYCATSSMY

RESULT 14

AP004829/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AP004829 304050 bp DNA linear BCT 02-JUL-2002

Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,

AP004829 BA0000033

AP004829.1 GI:21205117

Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.

Bacteria; Firmicutes; Bacillales; Staphylococcus.

1

Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,  
Nagai,Y., Iwama,N., Asano,K., Nishi,T., Kuroda,H., Cui,L.,  
Yamamoto,K. and Hiramatsu,K.

Genome and virulence determinants of high virulence  
community-acquired MRSA

JOURNAL Lancet 359 (9320), 1819-1827 (2002)  
MEDLINE 22040717  
PUBMED 12044378  
2 (bases 1 to 304050)  
AUTHORS Aoki, K., Oguchi, A., Nagai, Y., Asano, K., Iwama, N., Baba, T.,  
Kuroda, M., Hiramatsu, K. and Kikuchi, H.  
TITLE Direct Submission  
SUBMITTED (06-MAR-2002) Akio Oguchi, National Institute of  
Technology and Evaluation, Biotechnology Center; 2Chome 49-10  
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/  
Tel:81-3-3481-9423, Fax:81-3-3481-8424)  
FEATURES  
Location/Qualifiers  
1. .304050  
/organism="Staphylococcus aureus subsp. aureus MW2"  
/strain="MW2"  
/db\_xref="taxon:196620"  
/complement (629. .1072)  
/gene="MW1947"  
/complement (629. .1072)  
/gene="MW1947"  
/note="ORFID:MM1947  
hypothetical protein, similar to acetyltransferase  
[genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95812.1"  
/db\_xref="GI:21205118"  
/translation="MSITRLPNNSDPEKLNOCUKLYDGLPTNENDLKKRLKITT  
HDDYFLILLIKENKIKIGLSGCKMKPEKRAEYMRILAIVISEFRKKYGGKRLADS  
EERSKLNCALITLNSGNRNERLSAHKLISDNGVNTSGFTTKOL"  
/complement (1178. .1696)  
/gene="MW1948"  
/complement (1178. .1696)  
/note="ORFID:MM1948  
[genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical perotein"  
/protein\_id="BAB95813.1"  
/db\_xref="GI:21205119"  
/translation="MDNINKERKRVLTSTAITISFLNGLFSLAINGSNKNO  
CISLNLSDTFAVISWTTLGTALIGLAPSLYGFSPIDKSSFRASDWEEO  
RNWVKDMLMLPFLTSLILVIFIFQFAYSKLFTNTAIIIFIIILVMFPLILICYK  
ITWICRYFSKE"  
/complement (1932. .2522)  
/gene="MW1949"  
/complement (1932. .2522)  
/note="ORFID:MM1949  
hypothetical protein, similar to bacteriophage terminase  
small subunit [genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95814.1"  
/db\_xref="GI:21205120"  
/translation="MSELTAQAFVNEYIRTLNVTOSAIKAGYNANSAHTCRLLK  
KPHIKOYIOKOKIIDENVLAKELHLVLTNKTGDETEKEVVVVRGGYKKNPQSG  
KPOLVNEHVELEIETPKPSDRLLKARMLSKYHKLFPDKKDINGNVPIFINIGMDG  
DEGLDKAVKDSNDNPNHTIVDDIPLDEDYESTRFL"  
/complement (2519. .2734)  
/gene="MW1950"  
/complement (2519. .2734)  
/note="ORFID:MM1950  
[genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/product="truncated hypothetical protein"  
/protein\_id="BAB95815.1"  
/db\_xref="GI:21205121"  
/translation="WMTDDVNVNDHIIKNHLEMTIVDLATDKEFYIPDSLIIQGSYODI

gene  
2864. .3338  
/gene="truncated int"  
CDS 2864. .3211  
/gene="truncated int"  
/note="ORFID:MM1951  
truncated hypothetical protein, similar to integrase  
[genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95816.1"  
/db\_xref="GI:21205122"  
/translation="MKMSDIDPENNTISITTKTYVNPNNKKKKFQILPKPTSSIGKIS  
VDPNVATIKRDIKIVQNNWKNELINDHCNEKXTINVDINHTVTOHIOKXYALQI  
HTLRIAHRIKCH"  
3072. .3338  
/gene="truncated int"  
/note="ORFID:MM1952  
truncated hypothetical protein, similar to integrase  
[genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95817.1"  
/db\_xref="GI:21205123"  
/translation="MTTMYKLSMWIQTINRSTSLKNISTHSPRYTCALLIBSGV  
HIKEIOERLRHKDINTMVIYAKITNSYKDKASHKFSHOMEDVS"  
/complement (3433. .5049)  
/gene="groEL"  
/complement (3433. .5049)  
/note="ORFID:MM1953"  
/codon\_start=1  
/transl\_table=1  
/product="GroEL protein"  
/protein\_id="BAB95818.1"  
/db\_xref="GI:21205124"  
/translation="WYKQKPSFSDARQAMLRGVDOANAVKVTIGPKGNVLDKEFT  
APLITNDGVTLIAKEILEDYFENWGKLVQEVANKTNELAGCTTATYLAQMIOEG  
LNPVINSGANPVGLRQGIKRAVVAEALHENSQKVENKEIAQVGAISADEIGRYI  
SEAMERKVDGVITTEESNGLNTLEVEGMOFDRGQSPVWTSDDKVALEIRYI  
LVYDKKISFQDILPLLEQVOVSNRPIILVAEVEDDALTNVLTNMRGTFPAVAKA  
PBGDRKRMLEDIALITGAOYITDGLIDGASIDMGTSKVEVTMDTVNDG  
GBENSIDARVQSKSIOETESDPREKIOEHLAKLAVAYIKGASABETLKERKL  
RIEDALNSTRAVVEGIVAGGGLVNVYQVSEIABSDITGVNVLKALAPVRQ  
IENALIBSVLVERLKNAPGVPAATNENWMLBAGIVDPTKVTBSALOHAASA  
AMELITTEAVASIPERKNDQPNMGWPGHM"  
/complement (5125. .5409)  
/gene="groES"  
/complement (5125. .5409)  
/note="ORFID:MM1954"  
/gene="groES"  
/note="ORFID:MM1954"  
/codon\_start=1  
/transl\_table=1  
/product="GroES protein"  
/protein\_id="BAB95819.1"  
/db\_xref="GI:21205125"  
/translation="MLKPIGNRVITTEKBOQTGSGIVLTSAKESNEGVIYAVGT  
GRILNGTRVTPTEKGRDVVFEQYAGTEVKEDNETYLVLINEDLAVIE"  
5584. .6327  
/gene="MW1955"  
5584. .6327  
/note="ORFID:MM1955"  
/note="ORFID:MM1955"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAB95820.1"  
/db\_xref="GI:21205126"  
/translation="MTRLWASLTLVIIISQFLPLIIVKLPVQVSGSIELTKAVIY  
IOVLPLIATTIILINKIKNPTELEVBKPKYIIPALGFLVMIYQWVSIIV  
LTOIVGQOVSPNTEKLIITIAKPIEIFEYVIGPLIEEYVFRKVIPEGLPAIKGN  
RIVAFITATVSSLIPALAHNDKFIPIVFGGVIFSLAVYWTKRLAVPIIHHMLONG  
FVVIQPLNPEALKATEQANFIYHIFP"

```
gene complement (6352, .7611)
/ gene="MM1956"
CDS complement (6352, .7611)
/ note="MM1956"
/ note="ORFID:MM1956"
hypothetical protein, similar to SdH"
/ codon_start=1
/ trans1_table=1
/ protein_id="BAB5821.1"
/ db_xref="GI:21205127"
/ translation="MSYHFKKMLSTSLILSSSLGLATHTVEAKNDLNGKPTTN
LNHNTSPVSNSENNNTEGTPHESQTGNEGTSRDANPDSNNVKNPKDNNQNST
DSKPDNNQNSPNPKPDNDPKPKPDNDPKPKPDNDPKPKPDNDPKPKPDND
DKPKPDNDPKPDNDPKPDNDPKPDNDPKPDNDPKPDNDPKPDNDPKPDND
ASDGSNOGQMPKQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS
EATDEDIYNIIRKQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS
GEISGMDPKKPEKPKDSDKORSFEPHEKDFYVKKQEDNKKSASTAYKSKMLATYCS
MAYVSIIMLFVYKRNKKKNKESQR"
/ gene="MM1957"
7808, .8434
/ gene="MM1957"
7808, .8434
/ note="ORFID:MM1957"
/ codon_start=1
/ trans1_table=1
/ product="conserved hypothetical protein"
/ protein_id="BAB95822.1"
/ db_xref="GI:21205128"
/ translation="MGLFKDKKAMFPDNAMEERRSRITVNIKOSISISDELESMIHA
VKHPSFNSQSTRIVLLNDNNKFWMDTKALKWENGRNFEPEKIDPFKSHY
GTLFYEDDQVSGLOEMPNYDNFAIWTSTNMAHOFATWALTKTGIGSLQHYN
PILDEMTSNEFNPKSQKLIQAQPFQDIREAGEKTFNPFVDFIKK"
8795, .9580
/ gene="MM1958"
8795, .9580
/ gene="MM1958"
/ note="ORFID:MM1958"
/ codon_start=1
/ trans1_table=1

Query Match 4.2%; Score 44.6; DB 1; Length 304050;
Best Local Similarity 44.7%; Pred. No. 15;
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 635 AATATATTACTGTGAAGCAAGTAACCAATTTCAACTATAGTCGGATATAGTG 694
DB 262610 AAGGTGATACGTTGAATGTTGTTGAATATGTAACCAACATATAGTATPAATGTC 262551
C 695 AAGTATGTGTGAATTAAGATATCATCAGAGATCAACAATATCATGTTCAACTCATTTG 754
DB 262550 AATGTATGTAGGTTGAAGAAGGTTGAATGATTAATCTCATTTAAAGATTAATG 262491
QY 755 GTAACCTGTGAAGCCATAGCTTTATACGATGTTTGGCTTGAATTCGAATTTGTGTG 814
DB 262490 GCATGCGCTTATGATTTAAAGCTCTCTTTTATTTATTTAATTTTGGAGCAAGCC 262431
QY 815 ATGAAAAGCAAAAGGCTGCTGTTGAAGTTATCAAGATTTATGACATTAATCTTACTTTT 874
DB 262430 ATCCAAATTAATTAATAAATATCAATCAACAGCTGATATTAATAAATTTTAAAG 262371
QY 875 TGGACTGTGGTCTCGGATTAACCTCCAGCCCAAGAAATGATGCTCATTTGG 934
DB 262370 TTACATATCAGTTAAGATACAAATATTCAGACTAATTTTCAATTTGGCAAAATAT 262311
QY 935 CTAATATCATCAACTTTTATGCTCAATGAGCCAAACATTCGATGCGAAGAAAGTGAAG 994
DB 262310 CTTAAACATCAAAATTAATTAAGAAATTAATGATTTAACCATATCTAGATTAATAAAT 262251
QY 995 TTAGAGTTTGAGATGCTGACTTTG 1021
DB 262250 ATTGAATTTTATTAATTAATTTGTTAG 262224
```

```
RESULT 15
AP003136/c
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
genome, section 8/10.
ACCESSION
AP003136
VERSION
AP003136.2
KEYWORDS
Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315,
strain:N315) DNA.
SOURCE
Staphylococcus aureus subsp. aureus N315
Bacteria; Firmicutes; Bacillales; Staphylococcus.
ORGANISM
Staphylococcus aureus subsp. aureus N315
REFERENCE
1 Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iian, J., Ito, T., Kanamori, M.,
Matsumaru, H., Matuyama, A., Murakami, H., Hosoyama, A.,
Mizutani, U., Y., Takahashi, N., K., Sawano, T., Inoue, R., Kaito, C.,
Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
Kanehisa, M., Masahita, A., Oshita, K., Furuya, K., Yoshino, C.,
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
Lancet 357 (9264), 1225-1240 (2001)
JOURNAL
MEDLINE
21311952
PUBMED
11418146
TITLE
2 (bases 1 to 307750)
Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K.
and Kikuchi, H.
Direct Submision
Submitted (30-JAN-2001) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center, 2chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13701842.
COMMENT
Location/Qualifiers
1, .307750
/organism="Staphylococcus aureus subsp. aureus N315"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
297, .1907
/ gene="SA1853"
297, .1907
/ gene="SA1853"
/ note="ORFID:SA1853"
hypothetical protein, similar to DNA mismatch repair
protein MutS"
/ codon_start=1
/ trans1_table=1
/ protein_id="BAB43135.1"
/ db_xref="GI:13701843"
/ translation="MSTQTFLEIVIAIILLTSVIGIVGRVRSRQLKSMETLQFTI
SPLBAFIRPNSHFDEYKLYKEKFSHSVDDKTSDLNMAI FPKMNVNLPALIGEMK
LYACRGMISITNKSLLSLFNDAEFKRVYTHALIGTVPDPQITPVKRONIL
FLCPLPVSFAVIFINSQVGLFLMSGLFPIIISATLKRYEDLKSIFVASNVLK
OCYTSKIKIAGQPRVKNOPRTAARLHSLVAEVNDDIGAVTLVKILFMDVLF
HSIQSYTHNELKNCEDYIAELNHNALARRTRTLECYTEPQIDTDDIGVFSELT
HPLIADVAANDVLSQNILLTGNSNAGSTFKFSIAINILSAIQTVASAFVQPG
IVFTSMANADVLSGDSYFMALSKIKIIVEIPDOKIYCFIDEI FKGNTTERIAAS
ESVLSFLEHKSFRVIAATHDIELLELLKORYENHFVEIENNINHFDYKIKPKKAN
TRNAIEBKITSFPAKIVYERAKDNVSN"
complement (2222, .3247)
/ gene="SA1854"
complement (2222, .3247)
/ note="SA1854"
/ note="ORFID:SA1854"
hypothetical protein, similar to O-sialoglycoprotein
endopeptidase"
/ codon_start=1
/ trans1_table=1
/ protein_id="BAB43136.1"
/ db_xref="GI:13701844"
```





QY 815 ATGAAAAGCAAAAGCTTGGTGAAGTATCAAGAATTATTGACTAATCTTAACTT 874  
Db 256447 ATCAATTAAATATMAAAATATCAAAACACGTTGATATAGATTAAATTTAAAG 256388  
QY 875 TGGACTTGGGCTGGGATTAAGTCTCCAGCCAAAGAAATGATGCTGCTGCTTGG 934  
Db 256387 TTTACATATCAAGTTAAGATCAAAATATTCAGACTAAATTTCAATTTGGCAAAATAT 256328  
QY 935 CTAAATCATCAAACTTTATGCTCAATTTGAGCCAAATTCGATGCCAAGAAAGTGAAG 994  
Db 256327 CTTAAACATCAAAATTTATTAAGAAATAAATGTATTTAACCATATTTCTAGTAATAAAAAT 256268  
QY 995 TTAGAGTTTGAAGTGTGACTTTG 1021  
Db 256267 ATTGAATTTTAACTTAATTTGTTAG 256241

Search completed: January 27, 2003, 05:30:06  
Job time : 3762 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2003, 03:13:59 ; Search time 221 Seconds  
(without alignments)  
10882.960 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068  
Sequence: 1 atgtccaccacccaagaac.....gtgtcttgagaccattcctt 1068

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.8	4.1	14041	22	Internal control B
2	43	4.0	3346	18	Staphylococcus aur
3	42.2	4.0	50000	24	AMEPV genome fragm
4	41.6	3.9	3006	24	ABO69272
5	41.6	3.9	3024	24	ABO67866
6	41.6	3.9	40862	24	ABL34073
7	41.6	3.9	81905	24	ABO69244
8	41.6	3.9	82689	24	ABO67198
9	41.2	3.9	7491	21	AAV70156

10	41	3.8	9728	24	ABK40092	Human chemically p
11	41	3.8	9728	24	ABL33903	Human immune syste
12	41	3.8	15518	24	ABL70607	Chemically treated
13	41	3.8	15518	24	ABL34172	Human immune syste
14	41	3.8	15518	24	ABL34624	Human metastasis a
15	40.8	3.8	1537	22	AF58862	Porcine myoblastin
16	40.6	3.8	5407	22	AAJ34091	Human immune syste
17	40.4	3.8	7969	22	AAJ46679	Tumour suppressor
18	40.4	3.8	7969	24	ABO67037	Human angiogenesis
19	40.2	3.8	1572	20	AAJ61708	B. burgdorferi ant
20	40.2	3.8	8245	22	AAJ46447	Tumour suppressor
21	40	3.7	4530	22	AAJ24065	Yeast AOD9604-asso
22	40	3.7	17703	24	ABK39952	Human chemically p
23	39.8	3.7	6277	24	ABL33315	Human immune syste
24	39.8	3.7	13511	24	ABL32280	Human immune syste
25	39.6	3.7	2282	24	ABN95694	Gene #2192 used to
26	39.6	3.7	2287	16	AAJ00785	Human ataxin CDNA.
27	39.6	3.7	2287	16	AAJ02617	Human ataxin CDNA.
28	39.4	3.7	8842	24	ABJ33449	Human immune syste
29	39.4	3.7	8842	24	AAJ63335	Chemically pretrea
30	39.4	3.7	19659	24	ABL32766	Human immune syste
31	39.2	3.7	113515	24	ABL34174	Human immune syste
32	39	3.7	2324	20	AAJ20302	Borrelia burgdorfe
33	39	3.7	6973	24	ABL33778	Human immune syste
34	39	3.7	15767	24	ABL33206	Human immune syste
35	39	3.7	15767	24	ABL34552	Human metastasis a
36	39	3.7	580073	18	AAJ58840	Mycoplasma genital
37	38.8	3.6	5371	24	ABL33187	Human immune syste
38	38.8	3.6	6106	22	AAJ46429	Tumour suppressor
39	38.8	3.6	6106	24	ABK40031	Human chemically p
40	38.8	3.6	6106	24	ABL33472	Human immune syste
41	38.8	3.6	7011	24	ABK39940	Human chemically p
42	38.8	3.6	7011	24	ABL32546	Human immune syste
43	38.8	3.6	7067	24	AAJ59526	Long terminal repe
44	38.8	3.6	17967	24	ABL33015	Human immune syste
45	38.6	3.6	3923	22	ABA08191	Human ovarian and

#### ALIGNMENTS

RESULT 1  
AAH48024  
ID AAH48024 standard; DNA; 14041 BP.  
XX  
AC AAH48024;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Internal control B19c #1.  
XX  
KW Internal control; ss.  
XX  
OS Parvovirus.  
XX  
PN WO200146463-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 20-DEC-2000; 2000MO-EP12996.  
XX  
PR 22-DEC-1999; 99AT-0002170.  
XX  
PA (BAXT ) BAXTER AG.  
XX  
PI Zimmermann K, Turecek P, Schwarz H, Rieger M;  
XX WPI; 2001-408658/43.  
XX  
PT Internal standards useful for nucleic acid amplification assays,  
PT comprises a synthetic nucleic acid made by non-recombinant techniques  
XX

PS Example 1; Page 22-27; 30pp; English.

XX The present invention relates to methods for the preparation and use of  
CC internal controls for nucleic acid amplification assays. The internal  
CC controls comprise a synthetic nucleic acid made by non-recombinant  
CC techniques. The internal controls are useful for detecting nucleic acids  
CC in a sample such as blood, spinal fluid, semen, saliva, tears, cell  
CC culture fluid, recombinant cells, animal tissue or plant tissue, by a  
CC quantitative PCR assay, by adding the internal control to the sample,  
CC amplifying the nucleic acids in the sample and detecting the amplified  
CC products. The internal controls help in performing the nucleic acid  
CC amplification assay quickly, and inexpensively without sacrificing assay  
CC specificity or sensitivity. The present sequence is one such internal  
CC control, which was used in an example to illustrate the present  
CC invention.  
CC Note: the present sequence is the SEQ ID 8 shown in the sequence listing.  
CC This sequence differs from the SEQ ID 8 shown on page 13 of the  
CC disclosure (see AAH48038).

XX Sequence 14041 BP; 2151 A; 1354 C; 466 G; 2218 T; 7852 other;

Query Match

Best Local Similarity 4.1%; Score 43.8; DB 22; Length 14041;

Matches 97; Conservative 201; Mismatches 291; Indels 0; Gaps 0;

QY 464 CTAGTCTTTTCAAGTTTTCGATTGTATCAACAATTACTCAATCATCATCAGCTG 523

DB 10899 CHARTNTAHATGYAGHHNTAGNANMVTANSHAMATATAMTTARCTTTANAGSGM 10958

QY 524 CAGTGTATATCAAGCCTCTATCTCCACAATCTGGTGACCAAGTCAACAAGATATCA 583

DB 10959 NDSTATANAKATAMDVAGANDBSTTRSCNSDRNGTHVSDATASMNNDSTAAKRVANCN 11018

QY 584 CTCAAAATATAGAACCTTTTGGATTCAACAGTTGCTCCTCAACAGAGATATATTA 643

DB 11019 ARANATNRTKRGNTTAACTDANDTHRCARDRNTSCNSSTNGTHAMADRSTBT 11078

QY 644 ACTCTGAAGACAGTAAACCAATTCMAATCTATGCGATATAGTAAAGTATGT 703

DB 11079 ANMRSTVSAAGARATYCSHTNTNACNTRTHWSTNMRNCSARVAMAYASANDGTCSG 11138

QY 704 GTGAATTAAGATATATCTCAGAGATCAACAATATCTTCAACTATGGTACCTTG 763

DB 11139 NASRADDDTASMAHSAMWASTRACTDANDTHNSBCTDTCRTTHNSTDRASPSCRBD 11198

QY 764 ATAGCCATAGTATTAATCTGATGTTTGGCTTGAATTCGAATTTGGTGTGAAGAC 823

DB 11199 BYSMNDSTAAADTTTHNNSTDCRASDSCRBDYKAMTTAASAMSRDNDNTTACTD 11258

QY 824 AAAAGTTGCTGTAAGTATCAAGAATTTATGACTTAATTAAGTTTGGACTTGT 883

DB 11259 ANDBCTDTHDRNTCRNCGNSIVANDNGATVGNTRTHHTWSTHRTSARSMWARNT 11318

QY 884 TGGCTCTGGATTAATCTCCAGCCACAAGATGATTTGCTCATTTGGCTAAATCAT 943

DB 11319 ABTCMARSNHTHTHMTDSSHSWCMARBSNSIVTYANDSCCYRTMTHANDTHMTHD 11378

QY 944 CAACCTTTATGCTCAATTAAGCCAAATTCGATCGCAAGGAAGTGAAGTTAGATT 1003

DB 11379 SMMNDVSASASSTVWHAASNRHANDSTHCRSTASDSRCRBDYKAMTTADNTWKRCMC 11438

QY 1004 TGAGATGTGTGACTTTGCTAACAAGAAGTGAAGTTGCTGCTGT 1052

DB 11439 RSTWRACNTRNSRHNHTDNCASMSANDADTHTHTHRTSTSRSTVNCNT 11487

RESULT 2

AAV74457/c

XX AAV74457; standard; DNA; 3346 BP.

XX 16-MAR-1999 (first entry)

XX DE Staphylococcus aureus contig SEQ ID #146.

XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;

XX KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX KW skin infection; surgical wound infection; scalded skin syndrome;

XX KW toxic shock syndrome; ds.

XX OS Staphylococcus aureus.

XX FH Key

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

FT misc\_feature 2761..2820  
FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

FT misc\_feature /tag= b  
FT misc\_feature /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

QY 490 TATCAATATCTCCATATCATCATCTGACCTGATGATATCAAGCCTGATCTT 549  
 DB 2855 ATTATCTATCATATATAGATATATCAATTAANNNNNNNNNNNNNNNNNNNN 2796  
 QY 550 CCACATCTGTGTCACCAATGCAACAGATATCTCAAAAATATAGAACATTTGGAT 609  
 DB 2795 NNN 2736  
 QY 610 TCAACAGTGTGCTCTCAAGAGATATATATTAATCTGTAAAGCAAGTAAACCAAT 669  
 DB 2735 TCAACGACATGACCTTATATGCTAAATGTTAAACATGACTATTTATGTAATCTACT 2676  
 QY 670 TCAAACTACTATGTCGATATAGTAAAGTATGTGAAATTAAGATATCATCAGAGAT 729  
 DB 2675 TAAGATTGATATTTATGATTAAGATGATTTTAAGTTTAACTCA-ATATCCACATC 2618  
 QY 730 CAACATATCAATGTTCAACTGATGTTACTCTGATTAAGCC 770  
 DB 2617 TATTAATAATATGTTTAACTGATTAAGTCTTATACACC 2577  
 RESULT 3  
 ABL56201  
 ID ABL56201 standard; DNA; 50000 BP.  
 AC ABL56201;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE AMEPV genome fragment#3.  
 XX  
 KM AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;  
 XX  
 KM genetic deficiency disorder; ds.  
 XX  
 OS Ambacta moorei entomopoxvirus.  
 XX  
 PN WO200212526-A2.  
 PD 14-FEB-2002.  
 PF 10-AUG-2001; 2001WO-US25287.  
 XX  
 PR 10-AUG-2000; 2000US-224479P.  
 PR 14-SEP-2000; 2000US-0662254.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 XX Moyer RW, Li Y, Bawden AL;  
 DR WPI; 2002-227161/28.  
 XX  
 PT Novel recombinant entomopox virus vector useful for delivering  
 PT polynucleotide encoding protein to vertebrate cell, comprises  
 PT polynucleotide encoding protein operably linked with heterologous  
 PT promoter sequence -  
 XX  
 PS Disclosure; Page 175-201; 326pp; English.

CC production of sense or antisense nucleic acids. Vectors of the invention  
 CC provide for stable integration and expression of heterologous DNA in host  
 CC cells, and are adapted for accepting large heterologous polynucleotide  
 CC inserts which can be delivered in an infected or transformed cell and  
 CC expressed in a stable fraction. The current sequence represents a  
 CC fragment of the genome of the genus B entomopoxvirus from ambacta moorei  
 CC (AMEPV).  
 XX  
 SQ Sequence 50000 BP; 19754 A; 4704 C; 4365 G; 21177 T; 0 other;  
 Query Match 4.0%; Score 42.2; DB 24; Length 50000;  
 Best Local Similarity 46.2%; Pred. No. 2.5;  
 Matches 140; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
 QY 566 AAGTCACAGAGATATCATCTCAAAATATAGAACATTTGGATTCACAGTTGCTT 625  
 DB 33003 AATTATATCATTTATATATATTAATATATATTTCTGTTATTTGGATTTGAATATTTCT 33062  
 QY 626 CTCAGAGATATATTAATTAATCTGTAAAGCAAGTAAACCAATTTCAACTATGTCG 685  
 DB 33063 ATTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 33122  
 QY 686 GATATAGTGAAGTATGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 745  
 DB 33123 GCAATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 33182  
 QY 746 AACTGATGTCCTGATTAAGCAATGATTAATTAATTAATTAATTAATTAATTAATTA 805  
 DB 33183 ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 33242  
 QY 806 ATTTGTCATGAAGCAAGCAAGTTCGTTGAAGTTATCAAGAAATTTATGCTAATA 865  
 DB 33243 TAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 33302  
 QY 866 CTT 868  
 DB 33303 ATT 33305  
 RESULT 4  
 ABQ69272  
 ID ABQ69272 standard; DNA; 3006 BP.  
 AC ABQ69272;  
 XX  
 DT 29-AUG-2002 (first entry)  
 DE Listeria innocua DNA sequence #711.  
 XX  
 KM Antibacterial; Listeria; food contamination; mutational analysis;  
 XX  
 KM infection; ds.  
 XX  
 OS Listeria innocua.  
 XX  
 PN WO200228891-A2.  
 PD 11-APR-2002.  
 PF 04-OCT-2001; 2001WO-FR03061.  
 XX  
 PR 04-OCT-2000; 2000FR-0012697.  
 XX  
 PA (INSP ) INSR PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Kunst F, Glaser P;  
 DR WPI; 2002-332479/37.  
 XX  
 PT New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX

PS Claim 7; SEQ ID 2085; 180bp; French.

XX The present invention relates to nucleic acid sequences  
CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes  
CC and primers for identification and/or detection of *Listeria* (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
CC treating infections by *Listeria*, and are useful as immunogens in  
CC anti-*Listeria* vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3006 BP; 1108 A; 406 C; 581 G; 911 T; 0 other;

XX Query Match 3.9%; Score 41.6; DB 24; Length 3006;

XX Best Local Similarity 50.5%; Pred. No. 1.4;

XX Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

XX 43 TATCTTCCTTGGATGGAATGGAAGGCAAGAGTTAAATTGATCAATTGATCAAGAT 102

XX 475 TATATATTTTCAAAATTTGAAATGAAATGAAATACACAAATTTCTTAATGCGTTGAGCAA 534

XX 103 GTTTGGCAACTCAGGTTCCGATATATATGCAATATACCAATTTGATTGTTACAGT 162

XX 535 GTTGGCATATACAGATTTATGTCAGAGTCACGAATACATATATGATATATTACG 594

XX 163 GATGCTAGTCTTCAAGTCTTCCAGATGTTTCTCAACTGATAGCATTTCTTCATAT 222

XX 595 GCTGCATATCAAGAAATGCCCCAGAGTTTATCTATATATGCGTTATATATTTGTTT 654

XX 223 CTGTTTCTTGGGNGGT 242

XX 655 GGTGAATTTTGTCTGATTT 674

XX Db

XX RESULT 5

XX AB067866

XX ID AB067866 standard; DNA; 3024 BP.

XX AC AB067866;

XX DT 29-AUG-2002 (first entry)

XX DE *Listeria innocua* DNA sequence #668.

XX XX Antibacterial; *Listeria*; food contamination; mutational analysis;

XX XX infection; ds.

XX OS *Listeria innocua*.

XX PN WO200228891-A2.

XX PD 11-APR-2002.

XX PF 04-OCT-2001; 2001WO-FR03061.

XX PR 04-OCT-2000; 2000FR-0012697.

XX XX (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCT.

XX PI Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX DR

XX PT New genomic sequences from *Listeria* species, useful for detection,

XX PT treatment and prevention of infection, also related polypeptides,

XX PT antibodies and modulators

XX PS Claim 7; SEQ ID 679; 180bp; French.

XX The present invention relates to nucleic acid sequences

CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes  
CC and primers for identification and/or detection of *Listeria* (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
CC treating infections by *Listeria*, and are useful as immunogens in  
CC anti-*Listeria* vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3024 BP; 1114 A; 406 C; 590 G; 914 T; 0 other;

XX Query Match 3.9%; Score 41.6; DB 24; Length 3024;

XX Best Local Similarity 50.5%; Pred. No. 1.5;

XX Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

XX 43 TATCTTCCTTGGATGGAATGGAAGGCAAGAGTTAAATTGATCAATTGATCAAGAT 102

XX 493 TATATATTTTCAAAATTTGAAATGAAATGAAATACACAAATTTCTTAATGCGTTGAGCAA 552

XX 103 GTTTGGCAACTCAGGTTCCGATATATATGCAATATACCAATTTGATTGTTACAGT 162

XX 553 GTTGGCATATACAGATTTATGTCAGAGTCACGAATACATATATGATATATTACG 612

XX 163 GATGCTAGTCTTCAAGTCTTCCAGATGTTTCTCAACTGATAGCATTTCTTCATAT 222

XX 613 GCTGCATATCAAGAAATGCCCCAGAGTTTATCTATATATGCGTTATATATTTGTTT 672

XX 223 CTGTTTCTTGGGNGGT 242

XX 673 GGTGAATTTTGTCTGATTT 692

XX Db

XX RESULT 6

XX ABL34073

XX ID ABL34073 standard; DNA; 4082 BP.

XX AC ABL34073;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 2046.

XX XX Human, immune system disease; cytosine methylation; antiasthmatic;

XX KW antiarteriosclerotic; antiasthmatic; cytosine methylation; antiasthmatic;

XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX KW antineoplastic; antiarthritic; antidiabetic; antipsoriatic;

XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07537.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX XX (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX DR

XX

XX

XX

XX

XX

XX

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
PS Claim 1; SEQ ID NO 2046; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 40862 BP; 14301 A; 414 C; 7426 G; 18721 T; 0 other:  
Query Match 3.9%; Score 41.6; DB 24; Length 40862;  
Best Local Similarity 51.6%; Pred. No. 3.4;  
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 118 GTTCCGATATATATGATATACCAATTTGATTTTCACTGATGCTGATCTTCAA 177  
DB 18070 GGTGGCAAAATTTTATTTTATTTGAGTTTGTATTTGATGAGTTT 18129  
QY 178 AGTCTTCAGATGTTTCTCACTGATGATGATTTCTTCATATCTTTGGGT 237  
DB 18130 TGTGTGTAGAGATTTTATTTTATTTATTTATTTTATTTTATTTTGT 18189  
QY 238 GGTGTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 297  
DB 18190 TATTTGTTTGTGTTTGTGATGATGATGATGATGATGATGATGATGAT 18249  
QY 298 GTTT 301  
DB 18250 GTTT 18253

RESULT 7  
AB069244  
ID AB069244 standard; DNA; 81905 BP.  
XX  
AC AB069244;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria innocua DNA sequence #683.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
XX infection; ds.  
XX  
OS Listeria innocua.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaeser P;  
XX  
DR WPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators -  
XX  
PS Claim 5; SEQ ID 2057; 180pp; French.

XX  
CC The present invention relates to nucleic acid sequences  
CC (AB067198-AB071212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 81905 BP; 26162 A; 14249 C; 14844 G; 26650 T; 0 other:  
Query Match 3.9%; Score 41.6; DB 24; Length 81905;  
Best Local Similarity 50.5%; Pred. No. 4.2;  
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 43 TATCTTCTTGTGATGAAATGCGACGAGTTAATTGATCAATTTGATCAAGAT 102  
DB 24162 TATTAATTTTCAAAATTTGAAATGAAATGAAATGAAATGAAATGAAATG 24221  
QY 103 GTTTGCAACTCAGTTTCCGATATATATGCAATATACCAATTTGATTTACAGT 162  
DB 24222 GTTTGCAATATACGATTTATGTCATGAGTACGCAATACAGATTTAGATATACG 24281  
QY 163 GATGCTAGTCTTCAAAAGTCTTCCAGATGTTTCTCAACGATAGCATTTCTTCAAT 222  
DB 24282 GCTGCATATCAAGAGATGCGCCAGAGATTTATCTACTATATTTGCTATTAATTTGTTT 24341  
QY 223 CTGTTCTTGGGTGGTGT 242  
DB 24342 GGTGAATTTTGTCTGATTT 24361

RESULT 8  
AB067198  
ID AB067198 standard; DNA; 82689 BP.  
XX  
AC AB067198;  
XX  
DT 29-AUG-2002 (first entry)  
XX  
DE Listeria innocua plasmid DNA sequence.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
XX infection; ds.  
XX  
OS Listeria innocua.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaeser P;  
XX  
DR WPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators -  
XX  
PS Claim 5; SEQ ID 11; 180pp; French.

CC The present invention relates to nucleic acid sequences  
CC (ABQ67188-ABQ71212) from *Listeria* sp. The sequences are useful as probes  
CC and primers for identification and/or detection of *Listeria* (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
CC treating infections by *Listeria*, and are useful as immunogens in  
CC anti-*Listeria* vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIP0 at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 82689 BP; 26448 A; 14348 C; 14977 G; 26916 T; 0 other;  
Query Match 3.9%; Score 41.6; DB 24; Length 82689;  
Best Local Similarity 50.5%; Pred. No. 4.2;  
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
C 43 TATCTCTCTTGGATGAAATGCGAACGAGTTAAATTGATCATTTGATCAAGAT 102  
D 10877 TATATATATTTCAAAATTTGAAATGAAATACACACAACTTTAAATGCGTTGAGCAA 10936  
Qy 103 GTTTTGGCAACTCAGGTTTCCGATATATATCCAAATACCAGATTGATTGTTACAGT 162  
Db 10937 GTTTGGCAATATCAGATTAATGATGAGGTACGATACAGATTAATGATTAATTAACG 10996  
Qy 163 GATGCTAGTCTTCAAGCTTCCAGATGTTTCTCACTGATAGCATTTCTTCATAT 222  
Db 10997 GCTGCATATCAAGAAATGCCCCAGAGTTTATCTACTATATGCGTTATATATTTGTTT 11056  
Qy 223 CTGTGTTCTTTGGGTGTT 242  
Db 11057 GGTGAATTTTGTCTGATTT 11076  
RESULT 9  
ID AAA70156 standard; DNA; 7491 BP.  
XX AAA70156;  
AC  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:289.  
XX  
KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.  
XX  
XX Plasmodium falciparum.  
XX  
PN WO200025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
PE 05-NOV-1999; 99WO-US26796.  
XX  
PR 05-NOV-1998; 98US-0107131.  
XX  
PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 496-497; 577pp; English.

XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 7491 BP; 3252 A; 787 C; 964 G; 2488 T; 0 other;  
Query Match 3.9%; Score 41.2; DB 21; Length 7491;  
Best Local Similarity 46.0%; Pred. No. 2.5;  
Matches 139; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
Qy 580 ATCACTCAAAATATACACATTGTTGATTCACACTGTGCTCTCAAGAAATAT 639  
Db 373 ATTTATATATATAAATTAAGAAACTTTTGTGAACCTTTTGAATTTTGGGATAT 432  
Qy 640 ATTAACCTGTGAAGCAAGTAAACCAATTTCAACTACTATGTCGATATAGTAAAGT 699  
Db 433 ATATTTATATGATTAATATATTTTATATATATACATTAATGTTGATTAACAAA 492  
Qy 700 ATGTGAAATTAAGATATATCAGATCAACATACATATGTTCACTCATTTGTACC 759  
Db 493 AAGGTAAATAAAGAAATAGAAAGATTAAGAAATATGAAATATTAACATAATTAAT 552  
Qy 760 TCTGATTAAGCCATCGTTTATACGATGTTTGGCTTGAATCCATTTGTGTGATGAA 819  
Db 553 TATATATATATATTTTATGTTCAATATTTTATTTATTAATGATGAGAGAGAAAAA 612  
Qy 820 AAGCAAAAGTGTGCTGTGAAGTATCAAGATTTATGACTATACCTTTAGTTTGGAC 879  
Db 613 AAAAGAAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 672  
Qy 880 TT 881  
Db 673 GT 674  
RESULT 10  
ID ABK40092 standard; DNA; 9728 BP.  
XX ABK40092  
AC ABK40092;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human chemically pretreated gene sequence #87 strand 2.  
XX  
KM Human; de; bisulphite treatment; Cpg; DNA methylation; cancer; tumour;  
XX cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;  
XX UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
XX  
OS Homo sapiens.  
XX  
PN WO200202806-A2.  
XX  
PD 10-JAN-2002.



XX 29-JUN-2001; 2001WO-EP07470.  
 PF 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-154757/20.  
 PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligonucleotides,  
 PT useful for detecting cytosine methylation state of genes associated  
 PT with pharmacogenomics and for therapy of diseases e.g. cancer -  
 XX  
 PS Claim 1; SEQ ID No 174; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at  
 CC least 18 bases in length of a segment of the chemically pretreated DNA  
 CC of genes associated with pharmacogenomics according to one of the  
 CC sequences of the genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1  
 CC (NM\_000487), CYP3A3 (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2  
 CC (NM\_001979), OCIN (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360),  
 CC MRP (NM\_004996, NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898,  
 CC NM\_019899) and their complementary sequences, or a sequence (SI) chosen  
 CC from 87 sequences and their complements. The chemical pretreatment  
 CC is bisulfite treatment to convert cytosines (but not methyl-cytosines)  
 CC into uracils. Also included are an oligomer (II) in particular an  
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in  
 CC each case at least one base sequence having a length of 9 nucleotides  
 CC which hybridises to or is identical to a chemically pretreated DNA of  
 CC genes associated with pharmacogenomics and their complements, arranged in  
 CC an array for analysing diseases associated with the methylation state  
 CC (PPG) and/or detecting SNPs (single nucleotide polymorphisms)  
 CC of the 87 sequences. The oligomers may also be used as PCR primers.  
 CC The set of 87 nucleic acids and their complements is useful for diagnosis  
 CC and therapy of solid tumours and cancer. The present sequence  
 CC represents one the 87 DNA sequences or its complement.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC tcp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 other;

Query Match 3.8%; Score 41; DB 24; Length 9728;  
 Best Local Similarity 48.1%; Pred. No. 3;

Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 635 AATATATTAAGCTGTGAAGCAAGTAACCAATTTCAACTAGTCGATATAGTG 694  
 Db 8677 ATTTTATTAAGCTTTTATATATTTGGAGAAAGAAAAATATTTTGTGAATA 8736  
 QY 695 AAGATATGTGAATTAAGATATCATCAGAGATCAACAATAGTTCACATCTG 754  
 Db 8737 AAGTTTAATTAAGAAAAAGATATATATATATTTAGATAGTATTTATAG 8796  
 QY 755 GTACCTGTGATTAAGCCATACCTTATATCTGATGTTGGCTTGAATTCATTTG 814  
 Db 8797 GAAATGTAAATATGTGAGATGTATATTTTGAAGTTTATTTATATATATTTATTT 8856  
 QY 815 ATGAAAGCAAAAGGTGCTGTGAAGTATCAAGAAATTTTACATATCTTAGTT 874  
 Db 8857 GTAGTAGTAATTAAGATAGTAGCTTATTTATTTATATATATATATATAGATTTTGATT 8916  
 QY 875 T 875  
 Db 8917 T 8917

RESULT 11  
 ABL33903

ID ABL33903 standard; DNA; 9728 BP.

XX ABL33903;

AC 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1876.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianemic; cytosine; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; de.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JUN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX

PS Claim 1; SEQ ID NO 1876; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 other;

Query Match 3.8%; Score 41; DB 24; Length 9728;  
 Best Local Similarity 48.1%; Pred. No. 3;

Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 635 AATATATTAAGCTGTGAAGCAAGTAACCAATTTCAACTAGTCGATATAGTG 694  
 Db 8677 ATTTTATTAAGCTTTTATATATTTGGAGAAAGAAAAATATTTTGTGAATA 8736  
 QY 695 AAGATATGTGAATTAAGATATCATCAGAGATCAACAATAGTTCACATCTG 754  
 Db 8737 AAGTTTAATTAAGAAAAAGATATATATATATTTAGATAGTATTTATAG 8796  
 QY 755 GTACCTGTGATTAAGCCATACCTTATATCTGATGTTGGCTTGAATTCATTTG 814  
 Db 8797 GAAATGTAAATATGTGAGATGTATATTTTGAAGTTTATTTATATATTTATTT 8856  
 QY 815 ATGAAAGCAAAAGGTGCTGTGAAGTATCAAGAAATTTTACATATCTTAGTT 874  
 Db 8857 GTAGTAGTAATTAAGATAGTAGCTTATTTATTTATATATATATATAGATTTTGATT 8916  
 QY 875 T 875



Db 2630 AAAACCCCTTAAATTTAAAAATTTAACTCTTAATTAACCAATTAACATTAACCA 2571  
Qy 479 TTTTCGATTTGATCAACAATTTACTCCATCATCATCATGAGTGTGATTAACAAG 538  
Db 2570 TATTTTAAATTCATTTAATTTAATTTAATTAACATTAATTTAATTTAATTTA 2511  
Qy 539 CCTGTGATCTTCCCAATCTGGTGACCAAGTCAACAGATATCACTCAAAAATATAGAA 598  
Db 2510 ATCCCCATCTTTTAAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTA 2455  
Qy 599 CCATTTGGATTCACAGCTTGTGCTCTCAAGAGATATATTAATCTGTAAGCAAG 658  
Db 2454 -----TTCAAAATTTATTTCCCAATTAATAAAAAATTAATTAATAAAAAATC 2405  
Qy 659 GTAACCAATTTCAACTACTATGCGATATAGTAAGATGTGTAATTAAGATA 718  
Db 2404 GAAAAAATATTAATTAATTAACAAATCACTTAATTAATTTAATTTAATTTAATTTA 2345  
Qy 719 TCATCAGAGATCAACAATTAATGTTTC 745  
Db 2344 AAAACATTAATACTTCACTAATCTATTTC 2318

RESULT 14  
ABL34624/c  
ID ABL34624 standard; DNA: 15518 BP.  
XX  
AC ABL34624;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human metastasis associated gene SEQ ID NO: 177.  
XX  
KM Metastasis associated gene: cytostatic; gene therapy; cancer;  
XX cytosine methylation; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN M0200177376-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 06-APR-2001; 2001MO-EP03970.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-010922/01.  
XX  
XX  
XX New nucleic acid derived from chemically treated metastasis genes,  
XX PT useful for diagnosis of cancers by analysis of cytosine methylation,  
XX PT also for treatment -  
XX  
XX Claim 1; SEQ ID NO 177; 23pp + Sequence Listing; English.  
XX  
XX  
XX The present invention provides a number of human metastasis associated  
XX CC genes which are modified by cytosine methylation. The sequences can be  
XX CC used in the diagnosis and treatment of cancer. The present sequence is  
XX CC one of the genes of the invention.  
XX  
XX  
XX Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;  
XX  
XX  
XX Query Match. 3.8%; Score 41; DB 24; Length 15518;  
XX Best Local Similarity 48.3%; Pred. No. 3.5; Indels 14; Gaps 1;  
XX Matches 158; Conservative 0; Mismatches 155; Indels 14; Gaps 1;  
Qy 419 AAAAGTTGTTATGAACAATGTTATCCAGATGTTGCTTCTTAGTCTTTCAACAG 478

Db 2630 AAAACCCCTTAAATTTAAAAATTTAACTCTTAATTAACCAATTAACATTAACCA 2571  
Qy 479 TTTTCGATTTGATCAACAATTTACTCCATCATCATCATGAGTGTGATTAACAAG 538  
Db 2570 TATTTTAAATTCATTTAATTTAATTTAATTAACATTAATTTAATTTAATTTA 2511  
Qy 539 CCTGTGATCTTCCCAATCTGGTGACCAAGTCAACAGATATCACTCAAAAATATAGAA 598  
Db 2510 ATCCCCATCTTTTAAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTA 2455  
Qy 599 CCATTTGGATTCACAGCTTGTGCTCTCAAGAGATATATTAATCTGTAAGCAAG 658  
Db 2454 -----TTCAAAATTTATTTCCCAATTAATAAAAAATTAATTAATAAAAAATC 2405  
Qy 659 GTAACCAATTTCAACTACTATGCGATATAGTAAGATGTGTAATTAAGATA 718  
Db 2404 GAAAAAATATTAATTAATTAACAAATCACTTAATTAATTTAATTTAATTTAATTTA 2345  
Qy 719 TCATCAGAGATCAACAATTAATGTTTC 745  
Db 2344 AAAACATTAATACTTCACTAATCTATTTC 2318

RESULT 15  
AAF58662/c  
ID AAF58662 standard; DNA: 1537 BP.  
XX  
AC AAF58662;  
XX  
DT 27-APR-2001 (first entry)  
XX  
DE Porcine myostatin gene promoter.  
XX  
KM Porcine; pig; myostatin; promoter; cytostatic; anti-HIV; neuroprotective;  
XX immunomodulator; cancer; muscular dystrophy; spinal cord injury;  
XX KM neurodegenerative disorder; congestive obstructive pulmonary disease;  
XX KM COPD; amyotrophic lateral sclerosis; acquired immunodeficiency syndrome;  
XX AIDS; cachexia; ageing; ds.  
XX  
XX Sus scrofa.  
XX  
XX  
XX EP1072680-A1.  
XX  
XX  
XX 31-JUN-2001.  
XX  
XX 27-JUL-2000; 2000EP-0306396.  
XX  
XX 30-JUL-1999; 99US-0146540.  
XX  
XX  
XX (PF12 ) PFIZER PROD INC.  
XX  
XX  
XX Findly RC;  
XX  
XX  
XX WPI; 2001-170964/18.  
XX  
XX  
XX  
XX Novel myostatin gene promoter for regulating expression of heterologous  
XX PT genes in cells or animals, for identifying compounds that inhibit its  
XX PT activity or expression, which are useful for treating muscle disorders  
XX PT -  
XX  
XX  
XX Disclosure; Page 12-13; 24pp; English.  
XX  
XX  
XX The present sequence was used for comparison studies with a myostatin  
XX CC gene promoter sequence isolated from Mus musculus. Regulating the  
XX CC transcription activities or expression of the isolated promoter is useful  
XX CC for treating muscle associated disorders such as cancer, muscular  
XX CC dystrophy, spinal cord injury, neurodegenerative disorders, traumatic  
XX CC injury, congestive obstructive pulmonary disease (COPD), amyotrophic  
XX CC lateral sclerosis, acquired immunodeficiency syndrome (AIDS), cachexia  
XX CC or ageing. The promoter is useful for modulating the expression of the  
XX CC myosin gene for producing animal food products having increased muscle  
XX CC and protein content and reduced fat and cholesterol content, for

CC regulating expression of a heterologous gene in cells or animals, to  
CC engineer host cells, to screen for compounds that inactivate or inhibit  
CC its transcription and expression, for inhibiting its expression in cells  
CC for the promotion of muscle growth, for diagnostic evaluation, genetic  
CC testing and prognosis of a disease or disorder associated with myostatin  
CC expression.

XX

SQ Sequence 1537 BP; 519 A; 259 C; 236 G; 523 T; 0 other;

Query Match 3.8%; Score 40.8; DB 22; Length 1537;

Best Local Similarity 49.1%; Pred. No. 1.9; Mismatches 112; Indels 0; Gaps 0;

Matches 108; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 799 AATTCCAATTGTCGATGAAAAGCAAAAGGTCGTTGAGATTATCAGATTATTTG 858

Db 1284 AATTAAATCTATGTTAACAATAATATACTGCATTAATATTTATTTATTAG 1225

QY 859 ACTAATCTTTAGTTTGGACTTGTGGGCTCGGATTACTTCCAGCCACAGAT 918

P 1224 TGAACCTTAAATGTAATTTAAATAGTAATCTTTTACACTTCATTAATAAGTT 1165

QY 919 GGTATGCTCATTTGGCTAAATCATCAACTTTATGCTCAATTGAGCCACAATTGAT 978

Db 1164 GAAATGCTTTTCTCTCAAAAAATTAATGATTTTACATTATTTAGACAAACATTTGAG 1105

QY 979 GCCAAGGAAAGTGAAGTTAGATTTTGAGATGTTGACT 1018

Db 1104 GAAAAAAGAAATTATTTGCTGTTTAAATTAATTTTACT 1065

Search completed: January 27, 2003, 04:31:07  
Job time : 467 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 27, 2003, 04:18:00 ; Search time 52 Seconds  
(without alignments)  
6298.665 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068  
Sequence: 1 atgtccactcaacaaagac.....gtctcttgagaccatctcctt 1068

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.2	5.1	7218	1 US-08-232-463-14	Sequence 14, Appl
2	39.6	3.7	2287	1 US-08-222-619-1	Sequence 1, Appl
3	39.6	3.7	2287	1 US-08-221-767-23	Sequence 23, Appl
4	39.6	3.7	2287	5 PCT-US95-04075-1	Sequence 1, Appl
5	38	3.6	1103	1 US-08-553-633A-4	Sequence 4, Appl
6	37.2	3.5	2334	4 US-09-071-035-361	Sequence 361, App
7	36.6	3.4	19124	2 US-08-487-826B-13	Sequence 13, Appl
8	35.8	3.4	825	4 US-09-134-001C-1569	Sequence 1569, Ap
9	35.2	3.3	3706	4 US-08-913-159-9	Sequence 9, Appl
10	35.2	3.3	16950	4 US-09-453-702B-166	Sequence 166, App
11	35	3.3	1766	4 US-09-484-970B-132	Sequence 132, App
12	35	3.3	2728	3 US-08-836-402B-7	Sequence 7, Appl
13	35	3.3	5910	1 US-08-195-814-1	Sequence 1, Appl
14	34.6	3.2	3471	1 US-07-876-280-29	Sequence 29, Appl
15	34.6	3.2	3471	1 US-07-812-180A-1	Sequence 1, Appl
16	34.6	3.2	3471	1 US-08-315-650A-1	Sequence 1, Appl
17	34.6	3.2	3471	4 US-07-941-650A-1	Sequence 1, Appl
18	34.2	3.2	624	4 US-09-397-992A-3	Sequence 3, Appl
19	34.2	3.2	624	4 US-09-397-992A-6	Sequence 6, Appl
20	34	3.2	1750	4 US-09-276-531-34	Sequence 34, Appl
21	33.8	3.2	1968	4 US-09-298-724-1	Sequence 1, Appl
22	33.8	3.2	4182	4 US-07-973-257-1	Sequence 1, Appl
23	33.8	3.2	11443	4 US-08-961-527-49	Sequence 49, Appl
24	33.8	3.2	36159	4 US-09-749-586-3	Sequence 3, Appl
25	33.6	3.1	978	4 US-09-222-938A-45	Sequence 45, Appl
26	33.6	3.1	1071	4 US-08-887-534A-71	Sequence 71, Appl
27	33.6	3.1	1756	2 US-08-702-598-1	Sequence 1, Appl

C 28	33.4	3.1	985	4 US-08-842-306B-1	Sequence 1, Appl
C 29	33.4	3.1	985	4 US-08-838-973B-1	Sequence 1, Appl
C 30	33.4	3.1	985	4 US-08-771-212A-1	Sequence 1, Appl
C 31	33.4	3.1	19250	4 US-08-961-527-35	Sequence 35, Appl
C 32	33	3.1	723	4 US-09-134-001C-1263	Sequence 1263, Ap
C 33	33	3.1	1140	1 US-08-424-268-7	Sequence 7, Appl
C 34	33	3.1	1140	5 PCT-US93-10442-7	Sequence 7, Appl
C 35	33	3.1	1610	3 US-08-889-108-7	Sequence 7, Appl
C 36	33	3.1	1610	5 PCT-US94-10358-7	Sequence 7, Appl
C 37	33	3.1	1664	1 US-07-863-169A-6	Sequence 6, Appl
C 38	33	3.1	1664	2 US-08-429-964-6	Sequence 6, Appl
C 39	33	3.1	1664	3 US-07-935-087-6	Sequence 6, Appl
C 40	33	3.1	1664	5 PCT-US93-08062-6	Sequence 6, Appl
C 41	33	3.1	2160	4 US-08-188-275A-1	Sequence 1, Appl
C 42	33	3.1	2162	4 US-09-351-198-1	Sequence 1, Appl
C 43	33	3.1	2546	1 US-09-113-426-1	Sequence 1, Appl
C 44	33	3.1	2546	1 US-08-424-268-19	Sequence 19, Appl
C 45	33	3.1	2546	5 PCT-US93-10442-19	Sequence 19, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9pt-F1s  
US-08-232-463-14  
Query Match 5.1%; Score 54.2; DB 1; Length 7218;

Best Local Similarity 14.5%; Pred. No. 7.7e-05;  
Matches 83; Conservative 220; Mismatches 268; Indels 0; Gaps 0;

```

OY 142 ACCGAATTTGATGTTACAGATGATGATGTTCAAGTCTTCCAGATGTTTCTCACT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 ACAGAAATTAATTCGAGCTGCTGCGAGCTGAGAGAGCTTGCAGATTTTTTTTTT 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 202 GATAGCAATTTCTTCCATATCTGTTCTTTGGGTGCTCAAGATTTGATGAATCA 261
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1081 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1140
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 262 TTGGTTGCTGTGTTACTGGATTCATAGTTTGTTCCTCAATGCTCTGTCAAT 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1141 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 322 GGTTCGTTATGTTCCCAATATCTGCTCAAACTTTTATGCTCTCACAAAAT 381
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1201 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1280
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 382 GGTACTCAACAGATCTCCCTTTAGAAATGGCTCAAAAGTTGTTATGAACAAAT 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1261 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 442 GTTATTCAGATGTGCTCTCTAGTCTTTCACAGTTTTCGATGTATCAACAATTA 501
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1321 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1380
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 502 CTCATATCATCATCATGCTGATGATGATCAAGGCTCTGATCTTCCCAATCTGT 561
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1381 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1440
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 562 GACCAAGTCAACAGATATCTCAAAAATATAGAACATTTGGATCAACAGTTGT 621
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1441 AATCTCTATCTCTTAACTTGAATGATAGTAAATTAACAGTATGCTACATGCC 1500
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 622 GCCCTCAAGAGATATATTAATCTGTAAAGCAAGTAAACAATTTCAACTACTAT 681
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1501 GTTTTGAACCTGAATAGATGCGCTAGGAAGCGATGCTAGTCAATCAACCACT 1560
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 682 GTCGATATAGTGAAGATGTGTAATTA 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1561 TTCTATTTAGAAATATATGATTAATAAATA 1591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 2

US-08-222-619-1  
Sequence 1, Application US/08222619  
Patent No. 5652352

## GENERAL INFORMATION:

APPLICANT: Lichenstein, Henri  
APPLICANT: Lyons, David  
APPLICANT: Wurfel, Mark  
APPLICANT: Wright, Samuel  
TITLE OF INVENTION: Protein  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Center, Patent Operations/RRC  
STREET: 1840 DeHavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: U.S.  
ZIP: 91320-1789

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,619  
FILING DATE:  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2287 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA

NAME/KEY: CDS  
LOCATION: 318..2117  
FEATURE:

NAME/KEY: mat\_peptide  
LOCATION: 381..2114  
FEATURE:

NAME/KEY: sig\_peptide  
LOCATION: 318..380  
US-08-222-619-1

Query Match 3.7%; Score 39.6; DB 1; Length 2287;  
Best Local Similarity 47.6%; Pred. No. 0.32; Mismatches 129; Indels 0; Gaps 0;  
Matches 117; Conservative 0;

```

OY 237 TGGTGTCAAGATTTGGATGAATCATTTGTTGCTGTGTTACTGTGATTTGATAGTTT 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 TAGTATTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 297 TGTTCCTCAAGTCTCTGTCAATGTTCCGTTATGTTTCCCAATACTTGCTC 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 AATTCATTAAGCTTTAATTAATGAGCTACTCTTTATTTTGAACATCAATTTCTATCA 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 357 AAATTTTATTTGCTCCACCAATGTTACTCAACAGATCTCCCTTTAGAAATGGC 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 CTTTTCATTTTACTCCATATTTGAGGCTCAATAATCAATTTTATTTCTTTCTTT 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 417 TCAAAAGTTGTTTGAACAAATTTGTTATCCAGATGTGCTCTTCTTATGTTTTCAC 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 TGTAAATGTTGTTTCTACAAAGATGAACACTAAACTTACAGGTTTATTTTCTT 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 477 AGTTT 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 GTTTT 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 3

US-08-221-767-23  
Sequence 23, Application US/08221767  
Patent No. 6268212

## GENERAL INFORMATION:

APPLICANT: Simonet, William S.  
APPLICANT: Lichenstein, Henri S.  
APPLICANT: Lyons, David E.  
TITLE OF INVENTION: Tissue Specific Transgene Expression  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc., U.S. Patent Operations/MNO  
STREET: 1840 DeHavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/221,767  
FILING DATE:  
CLASSIFICATION: 800  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2287 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown

TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 318..2117  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 381..2114  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 318..380  
US-08-221-767-23

Query Match 3.7%; Score 39.6; DB 4; Length 2287;  
Best Local Similarity 47.6%; Pred. No. 0.32;  
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 237 TGGTGTCAAGAGTTGGATGATCATTTGGTGTGTTACTGGTGAATTCATAGTTT 296  
D 116 TAGATATACATTTGTTGTTGATGATATTATAGATTGTTGCTCACAAGAGTCT 175  
QY 297 TGTTCCTCAAGGCCCTGTGTCATGTTCCGTTATAGTTTCCACAAATCTGTGCTC 356  
Db 176 AAATCCATAAGCTTTATATATTCAGGCTACTCTTTATTTGAAAACCTCATTTCTATCA 235  
QY 357 AAATCTTTTATGTCCTCACCAGATGTAACAAGCATCTTCCCTTTAGAAATGGC 416  
Db 236 CTTTTTCTATTTTACTCCATATTTGAGGCTCATAAATCCATTTTATTTCTTTT 295  
QY 417 TCAAAAGTTGGTTATGACAAATGTTTATCCAGATGTTGCCCTCTTCTAGTTCTTTCAC 476  
Db 296 TGTAAATGTGGTTTCTACAAAGATGAAGTAACTAAACTTACAGGTTTATTTTCTT 355  
QY 477 AGTTT 482  
Db 356 GTTTT 361

## RESULT 4

PCT-US95-04075-1  
Sequence 1, Application PC/TUS9504075  
GENERAL INFORMATION:  
APPLICANT: AMGEN INC.  
TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like  
TITLE OF INVENTION: Protein  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Center, Patent Operations/RRC  
STREET: 1840 DeHavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: U.S.  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04075  
FILING DATE:  
CLASSIFICATION:  
INFORMATION FOR SEQ. ID NO. 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2287 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 318..2117  
FEATURE:

NAME/KEY: mat\_peptide  
LOCATION: 381..2114  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 318..380  
PCT-US95-04075-1

Query Match 3.7%; Score 39.6; DB 5; Length 2287;  
Best Local Similarity 47.6%; Pred. No. 0.32;  
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 237 TGGTGTCAAGAGTTGGATGATCATTTGGTGTGTTACTGGTGAATTCATAGTTT 296  
D 116 TAGATATACATTTGTTGTTGATGATATTATAGATTGTTGCTCACAAGAGTCT 175  
QY 297 TGTTCCTCAAGGCCCTGTGTCATGTTCCGTTATAGTTTCCACAAATCTGTGCTC 356  
Db 176 AAATCCATAAGCTTTATATATTCAGGCTACTCTTTATTTGAAAACCTCATTTCTATCA 235  
QY 357 AAATCTTTTATGTCCTCACCAGATGTAACAAGCATCTTCCCTTTAGAAATGGC 416  
Db 236 CTTTTTCTATTTTACTCCATATTTGAGGCTCATAAATCCATTTTATTTCTTTT 295  
QY 417 TCAAAAGTTGGTTATGACAAATGTTTATCCAGATGTTGCCCTCTTCTAGTTCTTTCAC 476  
Db 296 TGTAAATGTGGTTTCTACAAAGATGAAGTAACTAAACTTACAGGTTTATTTTCTT 355  
QY 477 AGTTT 482  
Db 356 GTTTT 361

## RESULT 5

US-08-553-633A-4  
Sequence 4, Application US/08553633A  
Patent No. 5821077  
GENERAL INFORMATION:  
APPLICANT: SALMOND, GEORGE PC  
APPLICANT: HOLDEN, MATTHEW TG  
APPLICANT: COX, ANTHONY RJ  
APPLICANT: THOMSON, NICHOLAS R  
APPLICANT: MCGOWAN, SIMON J  
TITLE OF INVENTION: PROCESS FOR ACTIVATING GENE EXPRESSION  
TITLE OF INVENTION: IN BACTERIA  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/553,633A  
FILING DATE: 01-Apr-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 1009-104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ. ID NO. 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1103 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Serratia  
US-08-553-633A-4

Query Match 3.6%; Score 38; DB 1; Length 1103;  
Best Local Similarity 43.6%; Pred. No. 0.67;  
Matches 170; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

598 ACATTTGGATTCACAGTGTGGCTCCCAAGAAATATTAATCTCTGTAACCA 657  
Db 330 ACCATATATGATTAATTAATCTTTATTAATGATCTGTTACTTCTTTAAAGA 389  
Qy 658 GGTAAACCAATTCACATCTATGTCGATATAGTAAAGTATGTGAATTAAGAT 717  
Db 390 GTAACTCAATTTCTGCGATGATGAAGATCTGCTGTTATAGCGTCGAAAGAT 449  
Qy 718 ATATCAGAGATCAACATGATGTTCAACTGTTGATCTCTGATAGCCATAGCTT 777  
Db 450 GTTGCCATGATCTAAGGAGATGATCATCTGATGTTACATTTGTTCTTCAAGAC 509  
Qy 778 TATATGATGTTTGGCTTGAATTCATTTGTGATGAAAGCAAAAGTGTCTGT 837  
Db 510 CATGATTAACATCTGGCATCTGACTATGCTAATGATGATAAATGATTTTGA 569  
Qy 838 GAATTAATCAAGATTTATGACTTAATTAATCTTGTGACTTGGCTGCGATTA 897  
Db 570 GATTTTATTAAGAACAGAAATGATTTACAATGTTGTAGACTACTCAGAAAA 629  
Qy 898 ACTCTCCAGCCAAAGATGTTATGCTCATTTGGCTAAATCATCAACTTTATGCT 957  
Db 630 GCATGAACATTAACCTTCGTTAAAGTAAACGGCCCTTGATGCTTCAAGT 689  
Qy 958 CAATTGAGCAACATTCGATGCAAGAA 987  
Db 690 GCATGATTAACACCGTGAACAGAAAGTA 719

## RESULT 6

US-09-071-035-361/C  
Sequence 361, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
CLASSIFICATION:  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 361:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2334 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-071-035-361

Query Match 3.5%; Score 37.2; DB 4; Length 2334;  
Best Local Similarity 55.4%; Pred. No. 1.4;  
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 189 TGTTCCTCACTGATGATGATTTCTTCATATCTGTTCTTTGGGTGTCAGAG 248  
Db 243 TGTTCCTCACTGATGATGATTTCTTCATATCTGTTCTTTGGGTGTCAGAG 184  
Qy 249 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 308  
Db 183 TGTTCCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
Qy 309 TGCCTCTGTC 318  
Db 123 AACAGTTGTC 114

## RESULT 7

US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhan  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121,001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13





Db 1081 AAGAAATACTATTCTATTATTAAGACTGGAGAAAAAAGTTAATCTTA 1026

## RESULT 10

US-09-453-702B-166/C  
; Sequence 166, Application US/09453702B  
; Patent No. 6365723

## GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.  
Burland, Nicole T.  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296,95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 166:

SEQUENCE CHARACTERISTICS:

LENGTH: 16950

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 166:

US-09-453-702B-166

## Query Match

Best Local Similarity 3.3%; Score 35.2; DB 4; Length 16950;

Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Db 342 ACAATACCTGGCTCAACTTTTATTTGCTCCACCAATGTAAGTCAAGCATCTTC 401  
Db 163 AGATTATATATGCTTCATTAATATATATCTCTCAACGATGCTGAATATATAC 104  
Db 402 CCTTTAGAAATGGCTCAAAAGTTGGTTATGAACAAATGTTATCCAGATGGCTC 461  
Db 103 CTATTAACAATCCGGTGTGAGATCTATTAACAATGAATGATTCTCTGCAATATATC 44  
Db 462 TTCTAGTT 469  
Db 43 TTTCAGTT 36

## RESULT 11

US-09-484-970B-132/C  
; Sequence 132, Application US/09484970B  
; Patent No. 6426186

## GENERAL INFORMATION:

APPLICANT: Jones, Karen A.

APPLICANT: Volkmuth, Wayne

APPLICANT: Walker, Michael G.

TITLE OF INVENTION: BONE REMODELING GENES

FILE REFERENCE: PB-0014 US

CURRENT APPLICATION NUMBER: US/09/484,970B

CURRENT FILING DATE: 2000-01-18

NUMBER OF SEQ ID NOS: 172

SOFTWARE: PERL Program

SEQ ID NO 132

LENGTH: 1766

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. 6426186 245065,1CB1

US-09-484-970B-132

## Query Match

Best Local Similarity 3.3%; Score 35; DB 4; Length 1766;

Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Db 609 TTCAACAGTGTGCTCTCAAGAGATATATTAATCTGTAAAGCAAGTAAACCAAT 668  
Db 334 TTCTAAAGTCTCTTCTAGCTCCAAAGCATGATTAATCTGTAGGCACTGAAAAAC 275  
Db 669 TTCAACTACTATGTCGATATAGTGAAGTATGTGAATTAAG 715  
Db 274 TGGAGATTATGAAGATGATCTCTCCAAATATATATATAAAG 228

## RESULT 12

US-08-836-402B-7/C

; Sequence 7, Application US/08836402B

; Patent No. 6063988

GENERAL INFORMATION:

APPLICANT: Rudiger Hain, Regina Fischer

TITLE OF INVENTION: DNA SEQUENCE AND ITS USE

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage

COMPUTER: Gateway 2000

OPERATING SYSTEM: DOS

SOFTWARE: Microsoft Windows 98

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/836,402B

FILING DATE: 02-SEP-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 44 40 200.7 (Germany)

FILING DATE: 10-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 9840-KGB

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2728 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
US-08-836-402B-7

Query Match 3.3% Score 35; DB 3; Length 2728;  
Best Local Similarity 50.9%; Pred. No. 5.4;

Matches 83; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 569 TCACAAGATATACCTCAAAAATATAGAACCATTTTGGATTCAAGTTGCTCTC 628

DB 1319 TCAGCATTTCTTGGTCAATGATGATGACGCTTCTGATCATGATTTTCTACCTGTA 1260

QY 629 AAGAGATATATTAATCTCTGTAAGCAAGTAAACCAATTTCAACTATGTCGAT 688

DB 1259 AAGAGATATAGATGACATGATGATTTTCCACGACTGATCATATATATCTTGAAG 1200

QY 689 ATAGTAAGATATGTAATTAAGTATATCATCATGATCA 731

DB 1199 ATGATCAACATGATTTTAACTGAGAAATTTAAAGGTTA 1157

RESULT 13

US-08-195-814-1/C  
Sequence 1, Application US/08195814

Patent No. 5547869

GENERAL INFORMATION:

APPLICANT: DUMAS, BRUNO; GERVAIS, MONICA;

APPLICANT: BERGION, MAX; JOURDAN, MIREITTE; JOUSSET,

APPLICANT: FRANCOISE XAVIERE

TITLE OF INVENTION: NOVEL PLASMIDS

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIERMAN AND MUSERLIAN

STREET: 600 THIRD AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08195,814

FILING DATE: 14-FEB-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/881,054

FILING DATE: 11-MAY-1992

APPLICATION NUMBER: 07/276,735

FILING DATE: 2-DEC-1988

ATTORNEY/AGENT INFORMATION:

NAME: CHARLES A. MUSERLIAN

REGISTRATION NUMBER: 19,683

REFERENCE/DOCKET NUMBER: 146,1029-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 661-8000

TELEFAX: (212) 661-8002

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5910

TYPE: NUCLEIC ACID

STRANDEDNESS: UNKNOWN

TOPOLOGY: UNKNOWN

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: DENSOVIRUS

STRAIN: DENSOVIRUS OF JUNONIA

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE: LARVAE

HAPLOTYPE:

TISSUE TYPE:  
CELL TYPE: SPODOPTERA LITTORALIS

CELL LINE:

ORGANELLE:

FEATURE:

LOCATION: 1

OTHER INFORMATION: N IS A OR C OR G OR T,

OTHER INFORMATION: WHEREIN N IS ZERO TO 50 NUCLEOTIDES IN LENGTH

FEATURE:

LOCATION: 1657

OTHER INFORMATION: M IS A OR C

FEATURE:

LOCATION: 5619

OTHER INFORMATION: Y IS C OR T

FEATURE:

LOCATION: 5910

OTHER INFORMATION: N IS A OR C OR G OR T,

OTHER INFORMATION: WHEREIN N IS ZERO TO 130 NUCLEOTIDES IN LENGTH

Query Match 3.3% Score 35; DB 1; Length 5910;  
Best Local Similarity 53.2%; Pred. No. 6.9;

Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 731 AACATACAAATGTTCACTGATGTAACCTGTGATTAAGCCATGTTATATGATGTTT 790

DB 3819 AATATATATGTCCTCCAGTACGATGATATGATGATGTCAGAAATTTGATGATGATTT 3760

QY 791 TGGCTTGAATTCATTTGTGTGATGAAGCAAGGTTGCTTTGAAGTTATCAAGA 850

DB 3759 TGTATGTGATCCTAAATAATGATGATATATACAGACGATCGATGACTTTGTAAAG 3700

QY 851 ATTATGACTAATGCTT 869

DB 3699 ATTAAATGCTATGCTTT 3681

RESULT 14

US-07-876-280-29  
Sequence 29, Application US/07876280

Patent No. 5262158

GENERAL INFORMATION:

APPLICANT: Payne, Jewel M.

APPLICANT: Cannon, Raymond J.C.

APPLICANT: Bagley, Angela L.

TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/876,280

FILING DATE: 19920430

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: M/S 104

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 3471 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: kumamotoensis  
INDIVIDUAL ISOLATE: PS50C  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC320) NRRL B-18769  
US-07-876-280-29

Query Match 3.2% Score 34.6; DB 1; Length 3471;  
Best Local Similarity 55.4%; Pred. No. 7.4; Indels 0; Gaps 0;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Q 416 CTCAAAAGTTGGTTATGACAAATGTTTATCCAGATGTCCTCTTCTAGTTCTTTCA 475  
D 491 CTAGAAAGTTGTTCTTATCAATTTATAGCTTAGATCTTAAGTTGTTCAATTC 550  
Q 476 CAGTTTCGATTGTATCAACAATTACTCCATCATCTCATCGCTGCAGTTGATATCA 535  
D 551 CATCTTTGCAGTATCCGACACGAACTATATTAGCAGTATATGACAGGCTGTGA 610

Q 536 A 536  
D 611 A 611

## RESULT 15

US-07-812-180A-1

; Sequence 1, Application US/07812180A  
; Patent No. 5366892  
; GENERAL INFORMATION:  
; APPLICANT: Foncerrada, Luis R  
; APPLICANT: Payne, Jewel M  
; APPLICANT: Sick, August J  
; TITLE OF INVENTION: No. 5366892el Coleopteran-Active Bacillus  
; TITLE OF INVENTION: thuringiensis isolate and a No. 5366892el Gene Encoding a  
; TITLE OF INVENTION: Coleopteran-Active Toxin  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Roman Saliwanchik  
; STREET: 2421 N.W. 41st Street, Ste A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/812,180A  
; FILING DATE: 19920102  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, Roman  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3471 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO

; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; STRAIN: kumamotoensis  
; INDIVIDUAL ISOLATE: PS50C  
; IMMEDIATE SOURCE:  
; LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF LUIS  
; LIBRARY: FONCERRADA  
; CLONE: 50C  
; US-07-812-180A-1

Query Match 3.2% Score 34.6; DB 1; Length 3471;  
Best Local Similarity 55.4%; Pred. No. 7.4; Indels 0; Gaps 0;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Q 416 CTCAAAAGTTGGTTATGACAAATGTTTATCCAGATGTCCTCTTCTAGTTCTTTCA 475  
D 491 CTAGAAAGTTGTTCTTATCAATTTATAGCTTAGATCTTAAGTTGTTCAATTC 550  
Q 476 CAGTTTCGATTGTATCAACAATTACTCCATCATCTCATCGCTGCAGTTGATATCA 535  
D 551 CATCTTTGCAGTATCCGACACGAACTATATTAGCAGTATATGACAGGCTGTGA 610

Q 536 A 536  
D 611 A 611

Search completed: January 27, 2003, 05:36:48  
Job time : 114 secs





Db 296 TGTAAAGTGGTTCTCAAGAGTAACTAAACTTACAGTTTATTTTCTT 355  
 Oy 477 AGTTT 482  
 Db 356 GTTTT 361

RESULT 5

US-09-783-590-6019  
 ; Sequence 6019 Application US/09783590  
 ; Patent No. US20020110850A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dillon, Patrick J.  
 ; APPLICANT: Haseltine, William A.  
 ; APPLICANT: Li, Haodong  
 ; APPLICANT: Rosen, Craig A.  
 ; APPLICANT: Ruden, Steven M.  
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
 ; PUBLICATION REFERENCE: PO-16.2C1  
 ; CURRENT APPLICATION NUMBER: US/09/783,590  
 ; CURRENT FILING DATE: 2000-02-15  
 ; PRIOR APPLICATION NUMBER: 08/420,856  
 ; PRIOR FILING DATE: 1995-04-12  
 ; PRIOR APPLICATION NUMBER: 08/346,731  
 ; PRIOR FILING DATE: 1994-11-21  
 ; NUMBER OF SEQ ID NOS: 12485  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6019  
 ; LENGTH: 490  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (10)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (58)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (119)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (134)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (140)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (162)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (203)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (205)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (262)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (266)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (285)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (294)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (299)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature

LOCATION: (301)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (305)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (322)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (327)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (347)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (358)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (365)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (370)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (372)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (376)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (388)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (397)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (401)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (403)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (424)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (426)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (427)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (430)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (431)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (433)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (434)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (449)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (450)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (452)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (461)

OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (462)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (467)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (472)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (473)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (476)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (489)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-6019

Query Match  
Best Local Similarity 3.6%; Score 38.2; DB 10; Length 490;  
Matches 70; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 739 AATGTCACATCTGCTGATGATAGCCATACGTTTACTGATGTTTGCTTTG 798  
DB 69 AGTAGCAGAGCTAGGATGATGATACAAAAGGTTATGGAAGTCTCTTG 128  
QY 799 AATTCACATTTGCTGATGAAAAGCAAAAGTGTCTGTTGAAGTATCAAAATTTATG 858  
DB 129 CATGTNAGTTTGTGAAGCAATGATGCTTTTGTGCTGATTAATCCAGATTCCTTGA 188

QY 859 ACTATA 865  
DB 189 AGGATA 195

## RESULT 6

US-09-834-975-317/c  
Sequence 317, Application US/09834975  
Patent No. US20020110815A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Hufel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
FILE REFERENCE: MRI-016B  
CURRENT APPLICATION NUMBER: US/09/834,975  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/197,538  
NUMBER OF SEQ ID NOS: 1046  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 317  
LENGTH: 536  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) (536)  
OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-317

Query Match  
Best Local Similarity 3.5%; Score 37.6; DB 10; Length 536;  
Matches 94; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
QY 334 GGTTCACAAATCTTGTCTCAACTTTTATTTGCTCCACCAATGTAATGCTCAACAA 393

DB 366 GTTTTNCCTTTTCTCTTTTNNAAAAANNTTTTAAAAAANNAATTCNAAATT 307  
QY 394 GCATCTTCCCTTTAGAAATGGCTCAAAAAGTTGTTATGACAAATGTTATCCAGAT 453  
DB 306 TTCCCGCCCTTTTAAACGGCCCAAAAGGATANTTAAAAAGTTTGGCCNCAAAAA 247  
QY 454 GTTGCCCTCTTCTAGTTCTTTACAGATTTTGGATTTGATCAACAATTTACCATCATCA 513  
DB 246 AAAAGCCCTTTCCTNNAAAAAGTTTGTATTTAAAGCAAAATGTTNCAATGAGA 187  
QY 514 TCATCAGCTGAGTGTATTC 534  
DB 186 NCAATTCAGATGCTAGATC 166

## RESULT 7

US-09-070-927A-180/c  
Sequence 180, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
APPLICANT: Patrick J. Dillon  
APPLICANT: Steven Barash  
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 180:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31517 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 180:  
US-09-070-927A-180

Query Match  
Best Local Similarity 3.5%; Score 37.2; DB 10; Length 31517;  
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 189 TGTTCCTCAACGATGCAATTTCTTCATATCTGTTCTTTGGGTGATGCAAG 248  
DB 24369 TGTTCCTCAACGATGATGTTATCTTGTCTGTGTGTTCACTTCGGCTTTTCTGC 24310  
QY 249 TTGGATGAATCATGTGCTGCTGTTACTGTGATTCATATGTTTGTTCCTCAAG 308



```
Db 24309 TGTATATGATGATGTTGCTTCTTACTTGTGTTTGTACTGTAATGTTTCGCAA 24250
QY 309 TGCCCTGTGC 318
Db 24249 AACAGTTGTC 24240

RESULT 8
US-09-946-807-1
; Sequence 1, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdotcitr, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
```

```
; OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1
Query Match 3.4%; Score 36.6; DB 9; Length 1503841;
Best Local Similarity 50.9%; Pred. No. 93;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGTTCCGTTATAGTTTCCCAACAATCTGTGCTCAAACTTTTATGTCCTCACCA 378
Db 1028986 AATTCATCAAGTTTATGTCCTGCGCCAGTCTTAAGCCCTCTTCTATGATTAACATCCA 1029045

QY 379 AATGTTCTCAACAAGATCTTCCCTTTGATGCTTAAAGTTGGTTATGAACAA 438
Db 1029046 GAAGATTAATCATCCAGCAGCTATCCGGGTACATTCGGCTGACTTGGTTAGTGGCTTAA 1029105

QY 439 AATGTTATCCAGATGTGCGCTTCTAGTCTTTGCACAGTTTCCGATGG 489
Db 1029106 GCAGTATATCCATGAGTCTTAACCTGGGTTTCTCAAACTTTTTCCTTTG 1029156

RESULT 9
US-09-795-668-1
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdotcitr, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=c/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
```

OTHER INFORMATION: n=a or g or c or t/u  
US-09-795-668-1

Query Match 3.4%; Score 36.6; DB 10; Length 1503841;  
Best Local Similarity 50.9%; Pred. No. 93;  
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGGTCCGTTTATGGTTCCCAATACTTGCTCAAACTTTTATATGTCTCAACA 378  
DB 1028986 AATTCTACAAGTTTATGTCTCTGCCCAAGTTCTTAAGCCCTTTCTAGATAACCAATCCA 1029045

QY 379 AATGTTCTCAACAGCATCTTCCCTTTTGAATGGCTCAAAAGTTGGTTAGAACAA 438  
DB 1029046 GAAGATAATCATCCAGAGCTATCCGGTACATTCGGCTGACTGTAGTTGGGTCTAA 1029105

QY 439 AATGTTATCCAGATGTTGCCTTTAGTCTTTTCAAGTTTCGATG 489  
DB 1029106 GCAGTATATCCATGAGTTCTTAAGTCCGGGTTTCTCAAACTTTTTCCTTTG 1029156

RESULT 10  
US-09-795-686-1  
Sequence 1, Application US/09795686  
Patent No. US20020094954A1  
GENERAL INFORMATION:  
APPLICANT: Stefansson, Hreinn  
APPLICANT: Steinhorsdottir, Valgerdur  
APPLICANT: Gulcher, Jeffrey R.  
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
FILE REFERENCE: 2345.2005-001  
CURRENT APPLICATION NUMBER: US/09/795.686  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 09/515,715  
PRIORITY FILING DATE: 2000-02-28  
NUMBER OF SEQ ID NOS: 1531  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1503841  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: r=g or a  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: y=c/u or c  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: m=a or c  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: k=g or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: s=g or c  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: w=a or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: d=a or g or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: b=a or c or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)  
OTHER INFORMATION: v=a or g or c  
NAME/KEY: misc\_feature  
LOCATION: (1)..(1531)

OTHER INFORMATION: n=a or g or c or t/u  
US-09-795-686-1

Query Match 3.4%; Score 36.6; DB 10; Length 1503841;  
Best Local Similarity 50.9%; Pred. No. 93;  
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGGTCCGTTTATGGTTCCCAATACTTGCTCAAACTTTTATATGTCTCAACA 378  
DB 1028986 AATTCTACAAGTTTATGTCTCTGCCCAAGTTCTTAAGCCCTTTCTAGATAACCAATCCA 1029045

QY 379 AATGTTCTCAACAGCATCTTCCCTTTTGAATGGCTCAAAAGTTGGTTAGAACAA 438  
DB 1029046 GAAGATAATCATCCAGAGCTATCCGGTACATTCGGCTGACTGTAGTTGGGTCTAA 1029105

QY 439 AATGTTATCCAGATGTTGCCTTTAGTCTTTTCAAGTTTCGATG 489  
DB 1029106 GCAGTATATCCATGAGTTCTTAAGTCCGGGTTTCTCAAACTTTTTCCTTTG 1029156

RESULT 11  
US-09-938-842A-4858/C  
Sequence 4858, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938.842A  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIORITY FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 4858  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4858

Query Match 3.4%; Score 35.8; DB 9; Length 2000;  
Best Local Similarity 48.7%; Pred. No. 17;  
Matches 97; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 669 TTCAACTACTANGTCGGATATAGTGAAGATGTAAGTATCATCAGAGA 728  
DB 1059 TTCACTTGTGGTCAATATACAGTAACATTTATTCACGGGTGACATTCGCATA 1000

QY 729 TCAACAATATACATGTTCAACTCATGTGATCTGTATAGCATACGTTTATGATGT 788  
DB 999 TCCACACACAGACAGAGTTATTAAGGATTTGATTTAGCCCTGATATTTGGTTCAA 940

QY 789 TTGGCTTTGAATTCATTTGTTGTGATGAAGCAAAAGTTGCTGTGAAGTTATCAA 848  
DB 939 ATATACGCTAATATACATTTCAATTTCTTTAAATAAATTTTATGTAATAGAGAACAA 880

QY 849 GAATTTATGACTAATCT 867  
DB 879 AATTTACCGAAATAATAT 861

RESULT 12  
US-09-818-512-3  
Sequence 3, Application US/09818512  
Patent No. US20020142416A1  
GENERAL INFORMATION:



DB 144 TGTAAATGAATGATGTCCTCTCTACTTGTGTTGACTGTGTAATGTTCTGCATA 85  
 QY 309 TGCCTCTGC 318  
 DB 84 AACAGTTGTC 75

RESULT 15

US-09-796-692-2970  
 ; Sequence 2970, Application US/09796692  
 ; Publication No. US20020198362A1

GENERAL INFORMATION:  
 ; APPLICANT: Gaiger, Alexander  
 ; APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

FILE REFERENCE: 2077,001200

CURRENT FILING DATE: 2001-03-01

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/190,479

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/200,545

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 60/200,303

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,779

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,999

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/202,084

PRIOR FILING DATE: 2000-05-04

PRIOR APPLICATION NUMBER: 60/206,201

PRIOR FILING DATE: 2000-05-22

PRIOR APPLICATION NUMBER: 60/218,950

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/222,903

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: 60/223,416

PRIOR FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: 60/223,378

PRIOR FILING DATE: 2000-08-07

NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2970

LENGTH: 371

TYPE: DNA

ORGANISM: Homo sapiens

US-09-796-692-2970

Query Match 3.3%; Score 35.2; DB 9; Length 371;  
 Best Local Similarity 50.6%; Pred. No. 13;  
 Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 582 CACTCAAAATATATGAAACATTGTTGATTCAACAGTTGTCCTCTCAAGGAAATATAT 641  
 DB 129 CACTGGAACCAACCAATTCAGGTCCTCAATGCTATTAATCTAGCAACAAACATTTA 188

QY 642 TAACTGTGTAAGCAAGGTAACCAATTTCAACTACTATGTCGATATAGTGAAGTAT 701  
 DB 189 CAGTTGTGCAAGAACAGATTCATTTCATTAACCAATTAACAAATTAACAAATTTTAT 248

QY 702 GTGTGAATTAAAGATATCATCAGATCAACAAATACATGTTCACT 749  
 DB 249 GTGTAAAGTAACTTACATCTTGGACTGTGAAAAATACITTTTAAT 296

Search completed: January 27, 2003, 07:48:41  
 Job time : 4938 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: January 27, 2003, 04:02:24 ; Search time 1718 Seconds  
(without alignments)  
10067.974 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068

Sequence: 1 atgtccaccacacaaagac.....gtgtcttgagaccatccct 1068

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

listing first 45 summaries

Database :

EST.\*  
1: em\_gesbda.\*  
2: em\_gesbhm.\*  
3: em\_geslin.\*  
4: em\_geslun.\*  
5: em\_geslov.\*  
6: em\_geslpl.\*  
7: em\_geslro.\*  
8: em\_geslro.\*  
9: gb\_gesl.\*  
10: gb\_gesl.\*  
11: gb\_gesl.\*  
12: gb\_gesl.\*  
13: gb\_gesl.\*  
14: gb\_gesl.\*  
15: em\_geslun.\*  
16: em\_geslun.\*  
17: gb\_gesl.\*  
18: em\_geslun.\*  
19: em\_geslun.\*  
20: em\_geslun.\*  
21: em\_geslun.\*  
22: em\_geslun.\*  
23: em\_geslun.\*  
24: em\_geslun.\*  
25: em\_geslun.\*  
26: em\_geslun.\*  
27: em\_geslun.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	51.8	4.9	1101	17	AL065414 Drosophila
C 2	51.8	4.9	1225	17	AL061171 Drosophila
C 3	44.6	4.2	1101	17	AL063921 Drosophila
C 4	44.2	4.1	887	17	AL060256 Drosophila
C 5	44	4.1	518	12	BG553567 df03c04.x
C 6	44	4.1	990	17	BG553567 df03c04.x

C 7	44	4.1	1101	17	CNS000EVL	AL069706 Drosophila
C 8	44	4.1	1101	17	CNS01219	AL101595 Drosophila
C 9	43.8	4.1	520	10	AM774971	AM774971 EST34122
C 10	43.6	4.1	537	13	BM276144	BM276144 PEST0A87
C 11	43.6	4.1	530	13	BM273621	BM273621 PEST0A85
C 12	43.6	4.1	533	13	BM276328	BM276328 PEST0A85
C 13	43.6	4.1	570	13	BM275241	BM275241 PEST0A86
C 14	43.6	4.1	747	17	CNS011RO	AL100640 Drosophila
C 15	43.6	4.1	1101	17	CNS0039G	AL063921 Drosophila
C 16	43.6	4.1	1101	17	CNS0002U	AL097152 Drosophila
C 17	43.4	4.1	649	12	BF298250	BF298250 015PDG01
C 18	43.4	4.1	870	17	CNS009XJ	AL054462 Drosophila
C 19	43.4	4.1	1001	17	CNS0155H	AL105023 Drosophila
C 20	43.4	4.1	1006	17	CNS010BY	AL098884 Drosophila
C 21	43	4.0	521	17	AZ336424	AZ336424 1M0066M10
C 22	43	4.0	628	17	AZ818307	AZ818307 2M0088R14
C 23	43	4.0	701	17	AZ316576	AZ316576 1M0034H16
C 24	42.8	4.0	467	17	CNS011DW	AL100142 Drosophila
C 25	42.8	4.0	526	13	BI814808	BI814808 PEST0A80
C 26	42.8	4.0	668	9	AL514901	AL514901 AL514901
C 27	42.8	4.0	734	17	CNS010MP	AL099163 Drosophila
C 28	42.8	4.0	869	17	CNS017Z2	AL108680 Drosophila
C 29	42.8	4.0	1121	9	AL554111	AL554111 AL554111
C 30	42.6	4.0	521	17	AQ810997	AQ810997 HS_5382_A
C 31	42.6	4.0	780	17	AQ842859	AQ842859 CPG1294B
C 32	42.6	4.0	1086	17	CNS012XX	AL102146 Drosophila
C 33	42.4	4.0	557	17	AQ505170	AQ505170 RPCR-11-2
C 34	42.4	4.0	978	17	AG129148	AG129148 Pan tlog1
C 35	42.4	4.0	996	17	CNS00F0H	AL071063 Drosophila
C 36	42.4	4.0	1101	17	CNS012TP	AL102007 Drosophila
C 37	42.2	4.0	409	17	AQ455933	AQ455933 HS_5151_B
C 38	42.2	4.0	658	17	B67167	B67167 CpG0046A Cp
C 39	42.2	4.0	1101	17	CNS017KX	AL108171 Drosophila
C 40	42.2	4.0	1191	14	BQ421837	BQ421837 AGENCOURT
C 41	42	3.9	839	12	BG400936	BG400936 602464975
C 42	42	3.9	1004	9	AL511881	AL511881 AL511881
C 43	41.6	3.9	803	17	AZ136058	AZ136058 SP_0171_A
C 44	41.6	3.9	811	17	CNS03YOO	AL266649 Tetraodon
C 45	41.6	3.9	963	17	CNS00X4L	AL054918 Drosophila

## ALIGNMENTS

RESULT 1  
CNS000D1/c 1101 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL065414.1 GI:4938827  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Oosagawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial



SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT			
Drosophila melanogaster	Drosophila melanogaster	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryogeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	1 (bases 1 to 1101)	Genoscope.	Direct Submission	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
Drosophila melanogaster	Drosophila melanogaster	The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .	Location/Qualifiers	1..1101	/organism="Drosophila melanogaster"	/db_xref="taxon:7227"	/clone="BACR08K10"	/clone_11b="RPCI-98"	/note="end : TET3"
BASE COUNT	201 a	64 c	131 g	202 t	503 others				
ORIGIN									
Query Match	4.2%	Score 44.6;	DB 17;	Length 1101;					
Best Local Smlarity	21.7%	Pred. 1.5;							
Matches	89;	Conservative 139;	Non.matches 183;	Indels	0;	Gaps 0;			
593	ATAGAACCATTGGATTCACAGCTGGCTCGCTCAAGAGAAATATATTAATCTCGTAA	652							
805	KRRBRKRAADRDADDDDAATTTTTRTTRDDDKMTDTWTRAAADRTDRDD	746							
653	ACCAAGTAAACCAATTTTCAACTACTATGTCGATATGATAAGATAGTGAATTA	712							
745	DDRBRAGTAGRMRRTWRKMRKRBDTRMDADADTDARDRRRRGGDADAKGKKTGK	686							
713	AGCATATCATCAGAGATCAACAATACATGTTCAACTATTGTAACCTCTGATAAGCAT	772							
685	RRRRBRATDRDADWADAAWMTTDTDDWDKRRRKGARRRRRTTARAAAMDMWTWK	626							
773	AGTTTATCTCATGTTTGGCTTGAATTCACATTTGCTGATGAAAAAGCAAAAGTTG	832							
625	AADMAKMDKTADRDRAADTWDARADADWAKARMRARRRARAARADRRWTTG	566							
833	CTGTGAAGTATCAAGAATTTATGACATTAATCTAGTTTGGACTGTGGGCTCG	892							
565	KTTATTTWTAAARAAWMAWMAATTTATTTTWTWTWTWTWTWTWTWTWTWTWT	506							
893	GATTAATCTCCAGCAACAAGAAATGATTTGCTCATTTGGCTTAATCATCAACTTT	952							
505	WTATAAATTAAMAAAAAAMATTTTTTTTTTTTTTWTAAAMWMTATTTTWTWT	446							
953	ATGCTCAATGAGCCACCAATTCGATGCCAAGAAAGTGAAGTGTG	1003							
445	AATTTTWTWTWTAAATTTTWTWTWTAAAAAAAAMWMTATATK	395							
RESULT 4	CNS00D5M	887 bp	DNA	linear	GSS 04-JUN-1999				
LOCUS	CNS00D5M	887 bp	DNA	linear	GSS 04-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # 899								

ACCESSION	BACR26H11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
VERSION	AL060256
KEYWORDS	AL060256.1 GI:4947618 GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryogeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 887) Genoscope.
REFERENCE	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
AUTHORS	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	source 1..887 /organism="Drosophila melanogaster" /db_xref="taxon.7227" /cd_name="BACR26H11" /clone_id="RPCI-98" /note="end : TE173"
BASE COUNT	242 a 109 c 62 g 257 t 217 others
ORIGIN	
Query Match	4.1%; Score 44.2; DB 17; Length 887;
Best Local Similarity	29.2%; Pred. No. 1.8;
Matches 202; Conservative 123; Mismatches 362; Indels 4; Gaps 1;	
Gy	AAAGGCAAGAAGTTAATTCATCTTGGATAAGAGATTGGCAACTCAGGTTT 121
Db	AAAAATTTATGCATTAATAATTAAGCTCTCCNCTAGGAAGTRCTGAATTCAGCTATCG 240
Gy	CCGATTAATAATGCAATATACCGAATTTGATTTGATTAAGTGATCTGTCAAATGC 181
Db	GTATAATAATATSWTCNNATWTSTTTTCTTCTTACTTAAKCTTTTWTWSKMAT 300
Gy	TTCAGATGTTTTCTCAACGTATAGCATTTTCTTCATATCTGTGTTCTTTGGGNGTG 241
Db	CTCTTTTKGCCNCTTTKKKGKGGSGGSSGCTCGCTTGCTGTGCGGGCKTKMGAT 360
Gy	TCAAGATTTGGATGATTCATTTGTCGTGCTGTGTTACTGCGATTTGCATAGTTGTTT 301
Db	KCTTKKKKGSGTTTNGTGTGKGGKKKKKTGTGNTTKGKGSBKTTCSNCGNCNTT 420
Gy	CCTCAAGCGCCGTGTCATGTTCCGTTTANGTTTCCCAACAATA----CTTGAGCTCA 357
Db	CSRSSSSSSSBTTTTTTTTTBSRBAITTTTWTWMACTTATCTBTTTSSBBYTT 480
Gy	AACCTTTATTGTCCTCACCAATGTTACTCAACAGCATCTTCCCTTTGAATTGGCT 417
Db	TMTWATTARVAAAATATCWACAYCCTTWYTAMTAAMTTTVAATATCTTTVCYAAAY 540
Gy	CAAAAGTTGGTATGAACAATTTGTTATCAAGATGTTGCCCTCTTCTAGTTCTTTCCA 477
Db	TCAATATATMACATTTTCMCATTTTWTAWAACVAAAATWTTTTCCTAAATATTTTMM 600







KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.  
FEATURES  
source  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_1id="BACN08024"  
/clone\_1lb="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : SP6"  
BASE COUNT 436 a 142 c 169 g 227 t 127 others  
ORIGIN  
Query Match 4.1%; Score 44; DB 17; Length 1101;  
Best Local Similarity 36.4%; Pred. No. 2.2;  
Matches 134; Conservative 42; Mismatches 192; Indels 0; Gaps 0;  
QY 127 TATATATCGAATATACGAATTGATTTGATGCTAGCTTCAAGCTTCCA 186  
DB 564 TTTTATATATATATACGGGTTTATATATATATATATATATATATATATA 505  
QY 187 GATGTTTTCACAGTATGAGATTTCTTCATATCTGTTCTGGGGTGTCAG 246  
DB 504 TTTTAT 445  
QY 247 AGTTTGATGATCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 306  
DB 444 ATTAT 385  
QY 307 AGTGCCTGTCGAATGCTCCGTTTATGCTTCCACAAATCTGCTCAACTTTTA 366  
DB 384 TTTTATCCCKKTTTATATATATATATATATATATATATATATATATATAT 325  
QY 367 TTGCTCACAAGAGTCAACAAGCATCTCCCTTTTGAATGGCTCAAAAGTT 426  
DB 324 TTTTAAATTTTWWTTTAAWMAWMAATTTTWTWMAWTTAAGTTATATTKW 265  
QY 427 GGTATGAACAATTTGTTATTCAGATGTGCTCTCTAGTCTTTCACAGTTTTCGA 486  
DB 264 TTTAAKWTATMCTTTTTCCTTTTATATATATATATATATATATATATAT 205  
QY 487 TTGTATCA 494  
DB 204 TTTWTTTA 197  
RESULT 9  
AM774971  
LOCUS AM774971 520 bp mRNA linear EST 07-SEP-2000  
DEFINITION ESTJ34122 KV3 Medicago truncatula cDNA clone pKV3-25E23, mRNA  
sequence.  
ACCESSION AM774971  
VERSION AM774971.1 GI:7718888  
KEYWORDS EST.  
SOURCE barrei medic.

ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 520)  
AUTHORS VandenBosch, K., Hurt, J., Moore, J., Beremand, P., Peng, H., Ellis, L.,  
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and  
Fraser, C.M.  
TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation  
JOURNAL Unpublished (1999)  
COMMENT Contact: VandenBosch K  
Department of Plant Biology  
University of Minnesota  
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA  
Tel: 612 624 2755  
Fax: 612 625 1738  
Email: kvandenb@chs.umn.edu  
Texas A&M EST name: MTEBK367K  
TIGR sequence name: MTEBK367K  
More information is available at:  
http://chryslie.tamu.edu/medicago  
Seq primer: Skmod (CTA GAA CTA gtc gat CC).  
FEATURES  
source  
1..520  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone\_1lb="pKV3-25E23"  
/clone\_1lb="KV3"  
/tissue\_type="Seedling roots"  
/dev\_stage="3 days post-inoculation with Sinorhizobium  
meliloti"  
/lab\_host="E. coli strain XL0LR"  
/note="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:  
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA  
was directionally ligated into the UniZap XR vector from  
Stratagene and packaged using Gigapack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-Zap phage using Ex-assist  
helper phage and propagated in XL0LR cells."  
BASE COUNT 188 a 92 c 110 g 130 t  
ORIGIN  
Query Match 4.1%; Score 43.8; DB 10; Length 520;  
Best Local Similarity 49.0%; Pred. No. 1.9;  
Matches 117; Conservative 0; Mismatches 122; Indels 0; Gaps 0;  
QY 620 TTGCTCTCAAGAGATATATTAATCTGTAAGCAAGTAAACCAATTTCAACTACT 679  
DB 247 TTGCTGCTCAAAAAGCAGCTGCAATGCTTAACAGCAAAACAAGCTGCTTCAAAAACA 306  
QY 680 ATGCGGATATAGTAAGATATGTAATTAAGATATATATATATATATATATAT 739  
DB 307 ATGTCACAGAAAAGAGAGATATATATATATATATATATATATATATATAT 366  
QY 740 ATGTTCAACTGATGTAAGTATGTAAGCAATATCTTTTACTGATGTTTGGCTTTGA 799  
DB 367 AAGTGTGTCGAAGAAGATCGAGATGTAAGAAAGTTTGAAGTCATGCTTTGATTTCAA 426  
QY 800 ATTGCAATTTGTGTGATGAAGCAAAAGTGTGCTGTTGAAGTTATTAAGAAATTTATG 858  
DB 427 ATATGAGAAATGATATGATCTGAAGTAACAAGTTAGAAAGATGCGAGACACAATAG 485  
RESULT 10  
BM276144  
LOCUS BM276144 527 bp mRNA linear EST 20-DEC-2001  
DEFINITION F8EST00a70h09.v1 Plasmodium falciptarum 3D7 gametocyte cDNA library  
Plasmodium falciptarum 3D7 cDNA 5', mRNA sequence.  
ACCESSION BM276144  
VERSION BM276144.1 GI:17969494  
KEYWORDS EST.

SOURCE Plasmodium falciparum 3D7.  
ORGANISM Plasmodium falciparum 3D7  
REFERENCE Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS 1 (bases 1 to 527)  
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,  
Matta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
Bowers, Y., Gibbons, M., Ritten, E., Bennett, J., Jentes, E., Ronko, I.,  
Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D.,  
Waterston, R., Wilson, R., and Sibley, D.  
TITLE MASHU Plasmodium EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: L. David Sibley  
MASHU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -400P from Gibco  
High quality sequence stop: 423.  
Location/Qualifiers  
1. 527  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_id="Plasmodium falciparum 3D7 gametocyte cDNA  
library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (Genehlog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:  
XhoI. The library was constructed by R. Haywood. cDNAs were  
synthesized from gametocyte poly(A) + RNA by oligo d(T)  
priming, size-selected and directionally cloned into the  
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR  
lambda vector (Stratagene). The primary library was mass  
excised as phagemid using the Exsist helper phage  
(Stratagene). Clones were mass excised using the Exsist  
helper phage (Stratagene), the phagemids were precipitated  
with PEG 8000 and extracted with phenol/chloroform.  
Phagemid DNA was electroporated into DH10B cells. Clone  
Availability: David Sibley, Washington University."

BASE COUNT 243 a 57 c 88 g 139 t  
ORIGIN  
Query Match 4.1%; Score 43.6; DB 13; Length 527;  
Best Local Similarity 48.8%; Pred. No. 2.2;  
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 807 TTTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATTGACTAATAC 866  
DB 55 TTTGTTAAGTGAATAAATGATATGATATATTAACAATTAACCATCAATTC 114

QY 867 TTTAGTTTGGACTTTGGGCTCGGATTACTCTCCACCAACAAGATGATATTC 926  
DB 115 ATCATATATGACCAAAATGATGCGGCTGCATTTTGGACCCACAATGATGTTATTGA 174

QY 927 TCATTGGCTAAATCATCAACTTTTATGCTCAATTGAGCCAAATTCGATGCAAGA 986  
DB 175 AGATAGCTGCATCACTCTCATCAGATCATGAATATATAGAAACAAATTAAGA 234

QY 987 AAGTGAAGTTAGATTGAGATGTGTTGACTTTGCTAACAGAAAGTTAAGATTGTC 1046  
DB 235 AAAACCGAAGCAGTAAAGAGTATGATTTGTATATGATTTAAAAAATTGA 294

QY 1047 TG 1048  
DB 295 TG 296

RESULT 11  
LOCUS BM273621 530 bp mRNA linear EST 20-DEC-2001  
DEFINITION PFE870a57c09.y1 Plasmodium falciparum 3D7 gametocyte cDNA library  
Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.  
ACCESSION BM273621  
VERSION BM273621.1 GI:17966918  
KEYWORDS EST.  
SOURCE Plasmodium falciparum 3D7.  
ORGANISM Plasmodium falciparum 3D7.  
REFERENCE Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS 1 (bases 1 to 530)  
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,  
Matta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
Bowers, Y., Gibbons, M., Ritten, E., Bennett, J., Jentes, E., Ronko, I.,  
Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D.,  
Waterston, R., Wilson, R., and Sibley, D.  
TITLE MASHU Plasmodium EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: L. David Sibley  
MASHU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -400P from Gibco  
High quality sequence stop: 422.  
Location/Qualifiers  
1. 530  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_id="Plasmodium falciparum 3D7 gametocyte cDNA  
library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (Genehlog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:  
XhoI. The library was constructed by R. Haywood. cDNAs were  
synthesized from gametocyte poly(A) + RNA by oligo d(T)  
priming, size-selected and directionally cloned into the  
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR  
lambda vector (Stratagene). The primary library was mass  
excised as phagemid using the Exsist helper phage  
(Stratagene). Clones were mass excised using the Exsist  
helper phage (Stratagene), the phagemids were precipitated  
with PEG 8000 and extracted with phenol/chloroform.  
Phagemid DNA was electroporated into DH10B cells. Clone  
Availability: David Sibley, Washington University."

BASE COUNT 246 a 57 c 88 g 139 t  
ORIGIN  
Query Match 4.1%; Score 43.6; DB 13; Length 530;  
Best Local Similarity 48.8%; Pred. No. 2.2;  
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 807 TTTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATTGACTAATAC 866  
DB 55 TTTGTTAAGTGAATAAATGATATGATATATTAACAATTAACCATCAATTC 114

QY 867 TTTAGTTTGGACTTTGGGCTCGGATTACTCTCCACCAACAAGATGATATTC 926  
DB 115 ATCATATATGACCAAAATGATGCGGCTGCATTTTGGACCCACAATGATGTTATTGA 174

QY 927 TCATTGGCTAAATCATCAACTTTTATGCTCAATTGAGCCAAATTCGATGCAAGA 986  
DB 175 AGATAGCTGCATCACTCTCATCAGATCATGAATATATAGAAACAAATTAAGA 234

QY 987 AAGTGAAGTTAGATTGAGATGTGTTGACTTTGCTAACAGAAAGTTAAGATTGTC 1046

Db	235	AAACCAAGACAGTAAAGAGCTAGTGGATTGTATATGATTATAAAAAAAAAATTGA	294
Qy	1047	TG	1048
Db	295	TG	296

RESULT 12					
BM276328					
LOCUS	BM276328	513 bp	mRNA	linear	EST 20-DEC-2001
DEFINITION	P53TGAa81d11.y1 Plasmodium falciparum 3D7 gametocyte cDNA library				
	Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.				

REFERENCE	ORGANISM	SOURCE	KEYWORDS	VERSION
1 (bases 1 to 533)	Plasmodium falciparum	Plasmodium falciparum 3D7	BM265328.1	GI:17969683
Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., P...	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium		EST.	

TITLE	Washu Plasmidium EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: L. David Sibley

Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence, stop: 424.

```

FEATURES
    source
        location/Qualifiers
            1..533
                /organism="Plasmodium falciparum 3D7"
                /db_xref="taxon:36329"
                /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                library"
                /dev_stage="gametocyte (stage III-V)"
                /lab_host="DH10B (Genescript, Invitrogen, Inc.)"
                /note="Vector: pBluescript SK plus; Site 1: EcoRI; Site
                2: XhoI. The library was constructed by R Haywood. cDNA
                synthesized from gametocyte poly(A)+ RNA by oligo d

```

BASE COUNT	249 a	57 c	88 g	139 t
ORIGIN				

Query Match	4.1 %	Score 43.6	DB 13	Length 533
Best Local Similarity	48.8 %	Pred. No. 2.2		
Matches 115; Conservative	0	Mismatches 124	Indels 0	Gaps 0

Qy	867	TTTACGTTTTGGACTCTGTGGGTCCTCGGATTAACCTCCAGCACAAGAATTGGTATTGC	922
Db	55	TTTGTTAACTGAATAAAAAATGATATCATATATAAAGCAATGAAAATCATCTTAATCC	114
Qy	807	TTTGTGTGATGAAAAGCAAAGGTGCTGTGAACTTATCAAGATTTATGACTAATAC	866

Db 115 ATCATATATGCACCAAAATGATGGCGCTCATTTTGCACCACCAACAAATGATGTATTGA 174

Qy 927 TCAATTGGCTAAATCATCAAACTTTTATGCTCAATTGAGCCAACTTGATGCCAAGA 986

Db 175 AGATACGTGCTCATCACTCTCATAGATCATGAATAATATGAGAAAAACAAATAATAAG 234

Qy 987 AAGTGAAGTGTAGAGTTTGGAGTGTGTACTTGTCTACACAGAACTTAAAGATTGTGC 104

Db 235 AAAACCCAAACACAGTAAAAAGAGTGTAGATTGTATTAATGATTTAAAAAAAATTGA 294

Qy 1047 TG 1048

Db 295 TG 296

RESULT 13	
BM275241	
LOCUS	570 bp mRNA linear EST 20-DEC-2001
DEFINITION	Pf18T0a66a06.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
ACCESSION	Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
VERSION	BM275241
KEYWORDS	BM275241.1 GI:17968565
SOURCE	EST.
ORGANISM	Plasmodium falciparum 3D7.
REFERENCE	Plasmodium falciparum 3D7.
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
	1 (bases 1 to 570)
	Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,

TITLE	Washu Plasmodium EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: L. David Sibley

Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 424.

```

FEATURES
    source
        location/Qualifiers
            1..570
                /organism="Plasmodium falciparum 3D7"
                /db_xref="taxon:36339"
                /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                library"
                /dev_stage="gametocyte (stage III-V)"
                /lab_host="DH10B (Genesys, Invitrogen, Inc.)"
                /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                XhoI; The library was constructed by R Haywood. cDNAs
                synthesized from gametocyte poly(A)+ RNA by oligo dT
                priming, size-selected and directionally cloned into

```

BASE COUNT	267 a	61 c	93 g	149 t
ORIGIN				

Query Match	4.18;	Score 43.6;	DB 13;	Length 570
-------------	-------	-------------	--------	------------

Query Match	4.1%	Score 43.6;	DB 17;	length 747;
Best Local Similarity	38.0%	Pred. No. 2.5;		
Matches	142;	Conservative	42;	Mismatches 188; Indels 2; Gaps 1;
Oy	100	GATGTTTTCGAACCTCAGGTTTCCGATATTAATGAAATATACCGAATTTGATTAC	159	
Db	478	GTTTTCGTTGCGRRATMTAGTTTTTTTTTTTTTTTTTTTTTTTATTTTGTCTTTTGT	419	
Oy	160	AGTATGCTAGCTCTTCAAGTCTTCAGATGTTTTCTCACTGATAGCATTTTCTCTCA	219	

363 TTTATTGCTTCACCAATGCTACTCAACAAGATCTTCCTTTTAGATTTGGCTCAAAA 422  
 Query Match 4.1%; Score 43.6; DB 17; Length 1101;  
 Best Local Similarity 14.3%; Pred. No.2.8;  
 Matches 79; Conservative 25; Mismatches 218; Indels 6; Gaps 1;

[illegible]

Search completed: January 27, 2003, 05:34:43  
Job time : 1730 secs